開始
NetApp Data Classification

NetApp
February 06, 2026

目錄

開始

了解NetApp Data Classification

NetApp Data Classification是NetApp Console的資料治理服務，它可以掃描您的企業內部和雲端資料來源以對應和分類資料並識別私人資訊。這可以幫助降低您的安全和合規風險，降低儲存成本，並協助您的資料遷移專案。

從 1.31 版開始，資料分類作為NetApp Console中的一項核心功能提供。無需額外付費。無需分類許可或訂閱。 + 如果您一直在使用舊版本 1.30 或更早版本，則該版本在您的訂閱到期之前可用。

NetApp Console

可以透過NetApp Console存取資料分類。

NetApp Console提供企業級跨本機和雲端環境的NetApp儲存和資料服務的集中管理。需要控制台才能存取和使用NetApp資料服務。作為管理介面，它使您能夠從一個介面管理許多儲存資源。控制台管理員可以控制企業內所有系統的儲存和服務的存取。

您不需要許可證或訂閱即可開始使用NetApp Console，並且只有當您需要在雲端部署控制台代理程式以確保與儲存系統或NetApp資料服務的連線時才需要付費。但是，一些可從控制台存取的NetApp資料服務是需要授權或基於訂閱的。

詳細了解"NetApp Console"。

特徵

資料分類使用人工智慧 (AI)、自然語言處理 (NLP) 和機器學習 (ML) 來理解其掃描的內容，以便提取實體並對內容進行相應的分類。這使得資料分類能夠提供以下功能領域。

"了解資料分類的用例" 。

保持合規

資料分類提供了多種工具，可以幫助您實現合規性。您可以使用資料分類來：

- 識別個人識別資訊 (PII)。
- 根據 GDPR、CCPA、PCI 和 HIPAA 隱私法規的要求識別廣泛的敏感個人資訊。
- 根據姓名或電子郵件地址回應資料主體存取請求 (DSAR)。

加強安全

資料分類可以識別可能被犯罪分子存取的資料。您可以使用資料分類來：

- 識別向整個組織或公眾公開的所有具有開放權限的文件和目錄（共用和資料夾）。
- 識別位於初始專用位置之外的敏感資料。
- 遵守資料保留政策。

- 使用*Policies*自動偵測新的安全性問題，以便安全人員可以立即採取行動。

優化儲存使用情況

資料分類提供可協助您降低儲存總擁有成本 (TCO) 的工具。您可以使用資料分類來：

- 透過識別重複或與業務無關的資料來提高儲存效率。
- 透過識別可以分層到較便宜的物件儲存的非活動資料來節省儲存成本。 "了解有關Cloud Volumes ONTAP系統分層的更多信息" 。 "了解有關本地ONTAP系統分層的更多信息" 。

支援的系統和資料來源

資料分類可以掃描和分析來自以下類型的系統和資料來源的結構化和非結構化資料：

系統

- Amazon FSx for NetApp ONTAP管理
- Azure NetApp Files
- Cloud Volumes ONTAP （部署在 AWS、Azure 或 GCP 中）
- 本地ONTAP集群
- StorageGRID
- Google Cloud NetApp Volumes

資料來源

- NetApp檔案分享
- 資料庫:
  - 亞馬遜關係型資料庫服務 (Amazon RDS)
  - MongoDB
  - MySQL
  - 甲骨文
  - PostgreSQL
  - SAP HANA
  - SQL 伺服器 (MSSQL)

資料分類支援 NFS 版本 3.x、4.0 和 4.1，以及 CIFS 版本 1.x、2.0、2.1 和 3.0。

成本

資料分類可以免費使用。無需分類許可或付費訂閱。

基礎設施成本

- 在雲端安裝資料分類需要部署雲端實例，這會導致部署雲端的雲端供應商收取費用。看為每個雲端提供者部署的執行個體類型 。如果您在本機系統上安裝資料分類，則無需付費。

- 資料分類要求您部署控制台代理程式。在許多情況下，由於您在控制台中使用其他儲存和服務，因此您已經擁有控制台代理程式。控制台代理執行個體會導致其部署所在的雲端提供者收取費用。查看 "為每個雲端提供者部署的執行個體類型"。如果您在本機系統上安裝控制台代理，則無需付費。

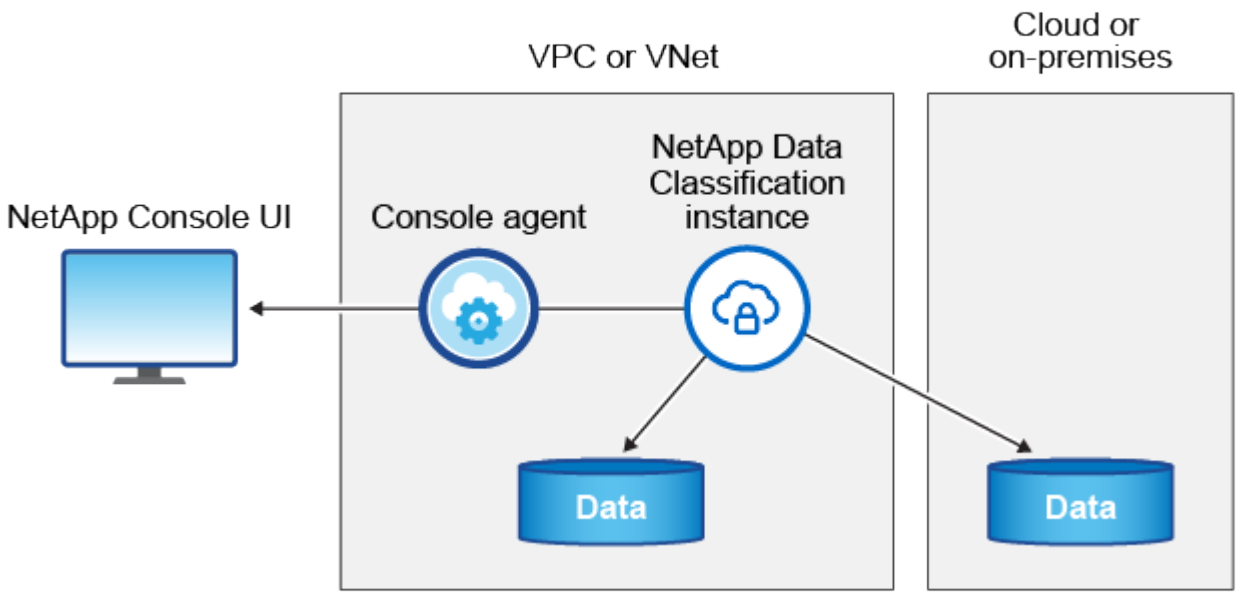
資料傳輸成本

資料傳輸成本取決於您的設定。如果資料分類實例和資料來源位於相同可用區和區域，則沒有資料傳輸成本。但是，如果資料來源（例如Cloud Volumes ONTAP系統）位於不同的可用區或區域，那麼您的雲端供應商將向您收取資料傳輸費用。請參閱以下連結以了解更多詳細資訊：

- "AWS：Amazon Elastic Compute Cloud (Amazon EC2) 定價"
- "Microsoft Azure：頻寬定價詳情"
- "Google Cloud：儲存傳輸服務定價"

資料分類實例

當您在雲端部署資料分類時，控制台會將執行個體部署在與控制台代理相同的子網路中。 "了解有關控制台代理的更多資訊。"



請注意預設實例的以下幾點：

- 在 AWS 中，資料分類在 "m6i.4xlarge 實例"帶有 500 GiB GP2 磁碟。作業系統映像是 Amazon Linux 2。在 AWS 中部署時，如果您要掃描少量數據，則可以選擇較小的執行個體大小。
- 在 Azure 中，資料分類在"Standard_D16s_v3 VM"帶有 500 GiB 磁碟。作業系統映像是 Ubuntu 22.04。
- 在 GCP 中，資料分類在"n2-standard-16 虛擬機"配備 500 GiB 標準持久性磁碟。作業系統映像是 Ubuntu 22.04。
- 在預設實例不可用的區域中，資料分類在備用實例上運行。"查看替代實例類型" 。
- 此實例名為 *CloudCompliance*，並帶有與之連接的產生的雜湊值（UUID）。例如：*CloudCompliance-16bb6564-38ad-4080-9a92-36f5fd2f71c7*
- 每個控制台代理程式僅部署一個資料分類實例。

您也可以在您的場所內的 Linux 主機上或您首選的雲端提供者的主機上部署資料分類。無論您選擇哪一種安裝方法，軟體的功能都完全相同。只要實例可以存取互聯網，資料分類軟體的升級就會自動進行。

實例應始終保持運行，因為資料分類會持續掃描資料。

在不同的實例類型上部署

查看實例類型的以下規範：

系統大小	規格	限制
特大號	32 個 CPU、128 GB RAM、1 TiB SSD	最多可掃描 5 億個文件。
大（預設）	16 個 CPU、64 GB RAM、500 GiB SSD	最多可掃描 2.5 億個文件。

在 Azure 或 GCP 中部署資料分類時，如果您想使用較小的實例類型，請發送電子郵件至 ng-contact-data-sense@netapp.com 尋求協助。

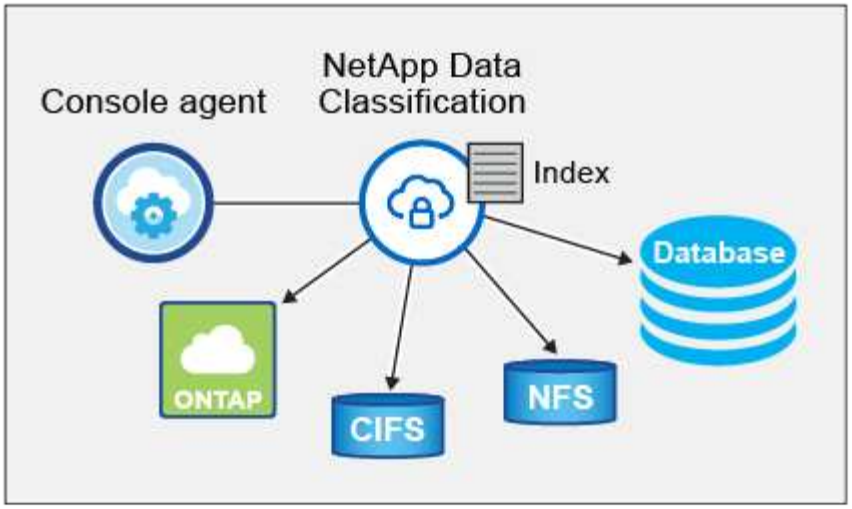
資料分類掃描的工作原理

從高層次來看，資料分類掃描的工作原理如下：

1. 您在控制台中部署資料分類實例。
2. 您可以在一個或多個資料來源上啟用進階映射（稱為「僅映射」掃描）或深層掃描（稱為「映射和分類」掃描）。
3. 資料分類使用人工智慧學習過程掃描資料。
4. 您可以使用提供的儀表板和報告工具來幫助您實現合規性和治理工作。

啟用資料分類並選擇要掃描的儲存庫（這些是磁碟區、資料庫模式或其他使用者資料）後，它會立即開始掃描資料以識別個人和敏感資料。在大多數情況下，您應該專注於掃描即時生產數據，而不是備份、鏡像或 DR 網站。然後，資料分類會對應您的組織數據，對每個檔案進行分類，並識別和提取資料中的實體和預定義模式。掃描結果是個人資訊、敏感個人資訊、資料類別和文件類型的索引。

資料分類透過安裝 NFS 和 CIFS 磁碟區像任何其他客戶端一樣連接到資料。 NFS 磁碟區會自動以唯讀方式訪問，而您需要提供 Active Directory 憑證來掃描 CIFS 磁碟區。



初步掃描後，資料分類將以循環方式持續掃描您的資料以偵測增量變化。這就是為什麼保持實例運行很重要。

您可以在磁碟區層級或資料庫模式層級啟用和停用掃描。

> 資料分類不會對其可以掃描的資料量施加限制。每個控制台代理程式支援掃描和顯示 500 TiB 的資料。要掃描超過 500 TiB 的數據，"安裝另一個控制台代理"然後"部署另一個資料分類實例"。 + 控制台 UI 顯示來自單一連接器的資料。有關查看來自多個控制台代理的資料的提示，請參閱"使用多個控制台代理"。

映射掃描和分類掃描之間有什麼區別

您可以在資料分類中進行兩種類型的掃描：

- 僅映射掃描僅提供資料的進階概覽，並在選定的資料來源上執行。僅映射掃描比映射和分類掃描花費的時間更少，因為它們不存取文件來查看其中的資料。您可能希望首先執行此操作來確定研究領域，然後對這些領域執行地圖和分類掃描。
- 地圖和分類掃描 為您的資料提供深層掃描。

有關映射掃描和分類掃描之間的差異的詳細信息，請參閱"映射和分類掃描之間有什麼區別？"。

資料分類所分類的信息

資料分類收集、索引並分配以下資料的類別：

- 關於文件的*標準元資料*：文件類型、大小、建立和修改日期等等。
- 個人資料：個人識別資訊 (PII)，例如電子郵件地址、身分證號碼或信用卡號，資料分類使用檔案中的特定單字、字串和模式進行識別。"了解有關個人資料的更多信息" 。
- 敏感個人資訊：《一般資料保護規範》（GDPR）和其他隱私法規定義的特殊類型的敏感個人資訊（SPII），例如健康資料、種族血統或政治觀點。"了解有關敏感個人資料的更多信息" 。
- 類別：資料分類將掃描的資料分為不同類型的類別。類別是基於 AI 對每個文件的內容和元資料的分析的主題。"了解有關類別的更多信息"。
- 名稱實體識別：資料分類使用人工智慧從文件中提取人們的自然名稱。"了解如何回應資料主體存取請求" 。

網路概述

資料分類可以在您選擇的任何地方部署單一伺服器或叢集：在雲端或本地端。伺服器透過標準協定連接到資料來源，並在 Elasticsearch 叢集中對結果進行索引，該叢集也部署在同一台伺服器上。這使得能夠支援多雲、跨雲端、私有雲和本地環境。

控制台使用安全性群組部署資料分類實例，該安全性群組啟用來自控制台代理程式的入站 HTTP 連線。

當您在 SaaS 模式下使用控制台時，與控制台的連線透過 HTTPS 提供，並且您的瀏覽器和資料分類實例之間發送的私人資料使用 TLS 1.2 進行端對端加密保護，這表示NetApp和第三方無法讀取它。

出站規則完全開放。需要網路存取來安裝和升級資料分類軟體以及發送使用情況指標。

如果您有嚴格的網路要求，"了解資料分類聯繫的端點" 。

存取NetApp Data Classification

您可以透過NetApp Console存取NetApp Data Classification。

若要登入控制台，您可以使用您的NetApp支援網站憑證，也可以使用您的電子郵件和密碼註冊NetApp Console登入。"了解有關登入控制台的更多信息" 。

特定任務需要特定的控制台使用者角色。"了解所有服務的控制台存取角色" 。

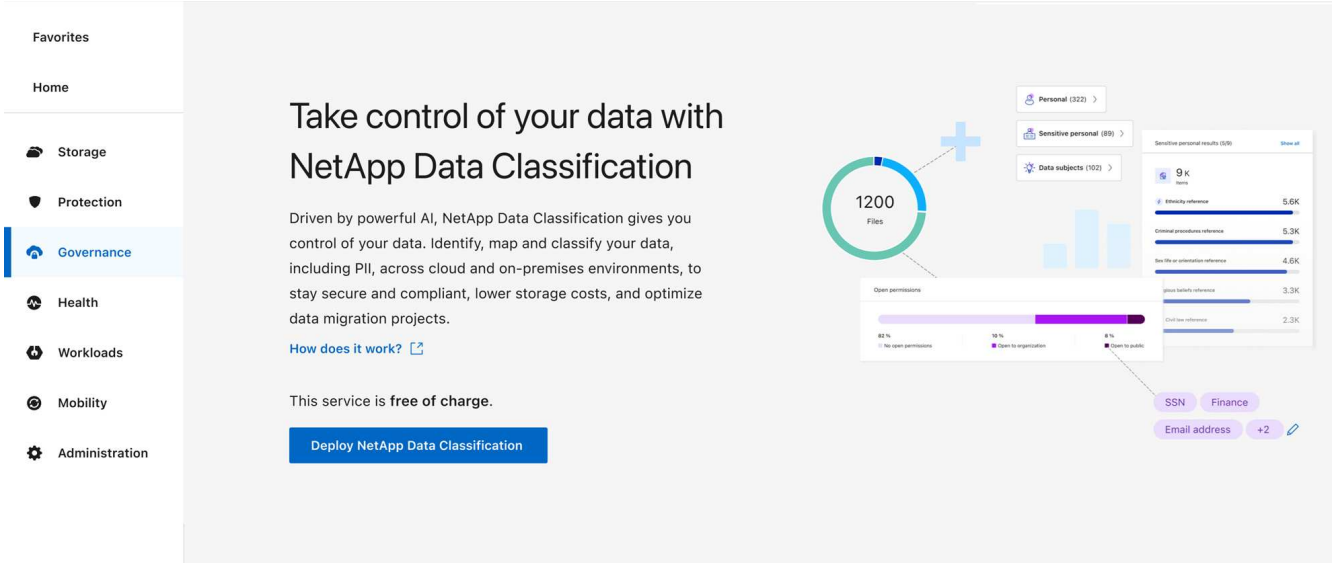
開始之前

- "您應該新增一個控制台代理程式。"
- "了解哪種資料分類部署樣式適合您的工作負載。"

步驟

1. 在 Web 瀏覽器中，導覽至"安慰"。
2. 登入控制台。
3. 從NetApp Console主頁中，選擇「治理」>「資料分類」 。
4. 如果這是您第一次造訪資料分類，則會出現登入頁面。

   選擇*在本機或雲端部署分類*以開始部署您的分類實例。有關詳細信息，請參閱"您應該使用哪種資料分類部署？"



否則，將出現資料分類儀表板。

部署資料分類

您應該使用哪種**NetApp Data Classification**部署？

您可以透過不同的方式部署NetApp Data Classification。了解哪種方法可以滿足您的需求。

資料分類可以透過以下方式部署：

- "使用控制台在雲端部署" 。控制台將資料分類執行個體部署在與控制台代理相同的雲端提供者網路中。
- "在可以存取網際網路的 Linux 主機上安裝" 。在您的網路中的 Linux 主機上或雲端中的 Linux 主機上安裝資料分類，前提是該主機可以存取網際網路。如果您希望使用同樣位於本機的資料分類實例來掃描本機ONTAP系統，則這種類型的安裝可能是一個不錯的選擇，儘管這不是必需的。
- "在沒有網際網路存取的本機站點的 Linux 主機上安裝"，也稱為_私人模式_。這種安裝類型使用安裝腳本，與控制台 SaaS 層沒有連接。

> BlueXP私有模式（傳統BlueXP介面）通常用於沒有網路連線的本機環境和安全雲端區域，其中包括 AWS Secret Cloud、AWS Top Secret Cloud 和 Azure IL6。NetApp繼續透過傳統的BlueXP介面支援這些環境。有關舊版BlueXP介面中的私有模式文檔，請參閱"BlueXP私人模式的 PDF 文檔"。

無論是在有網路存取的 Linux 主機上安裝，或是在沒有網路存取的 Linux 主機上進行本機安裝，都會使用安裝腳本。腳本首先檢查系統和環境是否符合先決條件。如果滿足先決條件，則安裝開始。如果您想獨立於執行資料分類安裝來驗證先決條件，您可以下載一個單獨的軟體包，該軟體包僅測試先決條件。

請參閱"檢查您的 Linux 主機是否已準備好安裝資料分類" 。

使用**NetApp Console**在雲端部署**NetApp Data Classification**

您可以使用NetApp Console在雲端部署NetApp Data Classification。控制台將資料分類執

行個體部署在與控制台代理理相同的雲端提供者網路中。

請注意，您還可以"在可以存取網際網路的 Linux 主機上安裝資料分類"。如果您希望使用同樣位於本機的資料分類實例來掃描本機ONTAP系統，則這種類型的安裝可能是一個不錯的選擇 - 但這不是必需的。無論您選擇哪一種安裝方法，軟體的功能都完全相同。

快速啟動

按照以下步驟快速開始，或向下捲動到其餘部分以獲取完整詳細資訊。


建立控制台代理

如果您還沒有控制台代理，請建立一個。看"在 AWS 中建立控制台代理"，"在 Azure 中建立控制台代理" ， 或者"在 GCP 中建立控制台代理"。

你也可以"在本機安裝控制台代理"在您網路中的 Linux 主機上，或是在雲端的 Linux 主機上。


先決條件

請確保您的環境能夠符合先決條件。這包括該實例的對外網際網路存取、Console agent 與 Data Classification 之間透過 443 埠的連線，以及其他要求。查看完整列表。


部署資料分類

啟動安裝精靈以在雲端部署資料分類實例。

建立控制台代理

如果您還沒有控制台代理，請在您的雲端提供者中建立控制台代理。看 "在 AWS 中建立控制台代理"或者 "在 Azure 中建立控制台代理"， 或者 "在 GCP 中建立控制台代理"。大多數情況下，您可能需要在嘗試啟動資料分類之前設定控制台代理，因為大多數 "控制台功能需要控制台代理"但有些情況下，你需要現在就設定一個。

在某些情況下，您必須使用部署在特定雲端提供者中的控制台代理：

- 在 AWS 中的Cloud Volumes ONTAP或Amazon FSx for ONTAP儲存桶中掃描資料時，您可以使用 AWS 中的控制台代理程式。
- 在 Azure 中的Cloud Volumes ONTAP或Azure NetApp Files中掃描資料時，您可以使用 Azure 中的控制台代理程式。
  - 對於Azure NetApp Files，它必須部署在與您要掃描的磁碟區相同的區域中。
- 在 GCP 中的Cloud Volumes ONTAP中掃描資料時，您可以使用 GCP 中的控制台代理程式。

使用任何這些雲端控制台代理時，都可以掃描本機ONTAP系統、 NetApp檔案共用和資料庫。

請注意，您也可以 "在本機安裝控制台代理"在網路或雲端的 Linux 主機上。一些計劃在本機安裝資料分類的使用者可能也會選擇在本機安裝控制台代理程式。

可能有些情況下你需要使用 "多個控制台代理"。

資料分類不會對其可以掃描的資料量施加限制。每個控制台代理程式支援掃描和顯示 500 TiB 的資料。要掃描超過 500 TiB 的數據，"安裝另一個控制台代理"然後"部署另一個資料分類實例"。 + 控制台 UI 顯示來自單一連接器的資料。有關查看來自多個控制台代理的資料的提示，請參閱"使用多個控制台代理"。

政府區域支持

當控制台代理部署在政府區域（AWS GovCloud、Azure Gov 或 Azure DoD）時，支援資料分類。以這種方式部署時，資料分類有以下限制：

"了解如何在政府區域部署控制台代理"。

先決條件

在雲端部署資料分類之前，請查看以下先決條件，以確保您具有受支援的配置。當您在雲端部署資料分類時，它與控制台代理程式位於同一子網路中。

啟用資料分類的出站互聯網訪問

資料分類需要出站網路存取。如果您的虛擬或實體網路使用代理伺服器進行網際網路訪問，請確保資料分類執行個體具有出站網際網路存取權限以聯絡下列端點。代理必須是非透明的。目前不支援透明代理。

根據您是在 AWS、Azure 還是 GCP 中部署資料分類，請查看下面的對應表格。

AWS 所需的終端節點

端點	目的
\ https://api.console.netapp.com	與控制台服務（包括NetApp帳戶）的通訊。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	與控制台網站通信，實現集中用戶身份驗證。
\ https://cloud-compliance-support-netapp.s3.us-west-2.amazonaws.com \ https://hub.docker.com \ https://auth.docker.io https://production.cloudflare.docker.com/ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/	提供對軟體映像、清單和範本的存取。
\ https://kinesis.us-east-1.amazonaws.com	使NetApp能夠從稽核記錄中串流資料。
\ https://cognito-idp.us-east-1.amazonaws.com \ https://cognito-identity.us-east-1.amazonaws.com \ https://user-feedback-store-prod.s3.us-west-2.amazonaws.com \ https://customer-data-production.s3.us-west-2.amazonaws.com	啟用資料分類來存取和下載清單和模板，以及發送日誌和指標。

Azure 所需的終點

端點	目的
\ https://api.console.netapp.com	與控制台服務（包括NetApp帳戶）的通訊。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	與控制台網站通信，實現集中用戶身份驗證。
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https https://dseasb33srnrn.cloudfront.net/ // https://production.cloudflare.docker.com/	提供對軟體映像、清單、範本的存取以及發送日誌和指標。
\ https://support.compliance.api.console.netapp.com/	使NetApp能夠從稽核記錄中串流資料。

GCP 所需的端點

端點	目的
\ https://api.console.netapp.com	與控制台服務（包括NetApp帳戶）的通訊。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	與控制台網站通信，實現集中用戶身份驗證。

端點	目的
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https https://dseasb33srnrn.cloudfront.net/ // https://production.cloudflare.docker.com/	提供對軟體映像、清單、範本的存取以及發送日誌和指標。
\ https://support.compliance.api.console.netapp.com/	使NetApp能夠從稽核記錄中串流資料。

確保資料分類具有所需的權限

確保資料分類具有部署資源和為資料分類實例建立安全性群組的權限。

- "Google Cloud 權限"
- "AWS 權限"
- "Azure 權限"

確保控制台代理可以存取資料分類

確保控制台代理程式和資料分類實例之間的連線。控制台代理程式的安全性群組必須允許透過連接埠 443 進出資料分類實例的入站和出站流量。此連線支援部署資料分類實例，並允許您查看「合規性和治理」標籤中的資訊。 AWS 和 Azure 的政府區域支援資料分類。

AWS 和 AWS GovCloud 部署需要額外的入站和出站安全群組規則。看 "AWS 中的控制台代理程式規則"了解詳情。

Azure 和 Azure 政府部署需要額外的入站和出站安全群組規則。看 "Azure 中的控制台代理程式規則"了解詳情。

確保資料分類能夠持續運行

資料分類實例需要保持開啟狀態以持續掃描您的資料。

確保 **Web** 瀏覽器連接到資料分類

啟用資料分類後，請確保使用者從與資料分類實例有連接的主機存取控制台介面。

資料分類實例使用私人 IP 位址來確保索引資料無法被網際網路存取。因此，您用來存取控制台的 Web 瀏覽器必須連線到該私人 IP 位址。此連線可以來自與雲端提供者的直接連線（例如 VPN），也可以來自與資料分類執行個體位於同一網路內的主機。

檢查您的 **vCPU** 限制

確保您的雲端提供者的 vCPU 限制允許部署具有必要數量的核心的執行個體。您需要驗證控制台運作區域中相關執行個體系列的 vCPU 限制。"查看所需的實例類型" 。

有關 vCPU 限制的更多詳細信息，請參閱以下連結：

- "AWS 文件：Amazon EC2 服務配額"
- "Azure 文件：虛擬機器 vCPU 配額"

- "Google Cloud 文件：資源配額"

在雲端部署資料分類

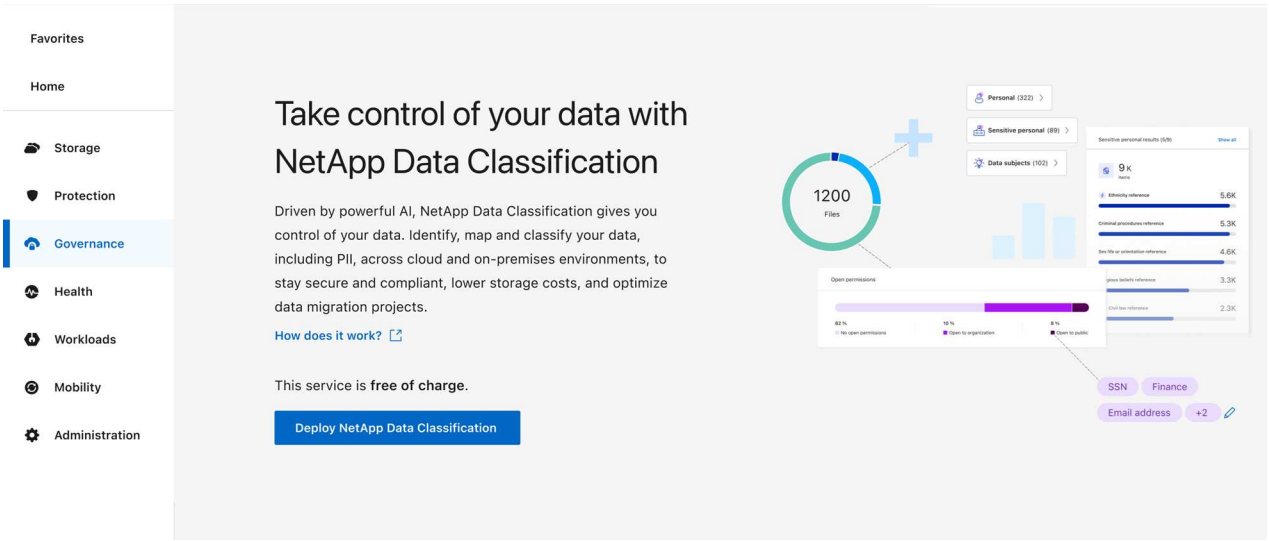
請依照下列步驟在雲端部署資料分類實例。控制台代理程式將在雲端部署實例，然後在該實例上安裝資料分類軟體。

在預設實例類型不可用的區域中，資料分類在"備用實例類型"。

在 **AWS** 中部署

步驟

1. 從資料分類主頁中，選擇*在本地或雲端部署分類*。

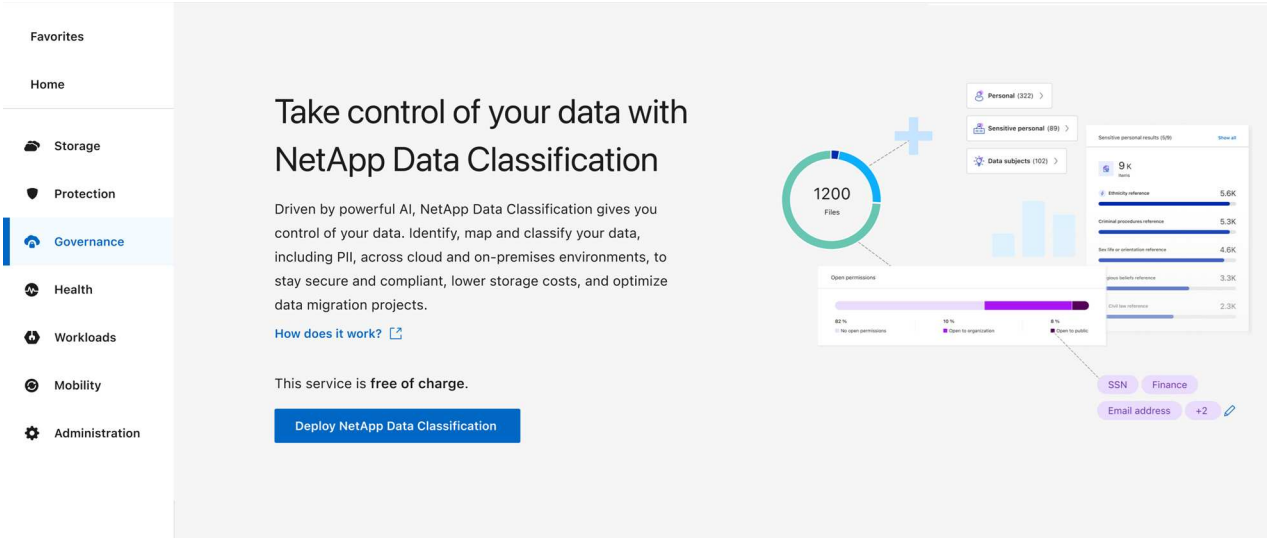


2. 從「安裝」頁面，選擇「部署」>「部署」以使用「大型」實例大小並啟動雲端部署精靈。
3. 精靈在執行部署步驟時會顯示進度。當需要輸入或遇到問題時，系統會提示您。
4. 當實例部署完畢並安裝資料分類後，選擇「繼續設定」進入「設定」頁面。

在 **Azure** 中部署

步驟

1. 從資料分類主頁中，選擇*在本地或雲端部署分類*。

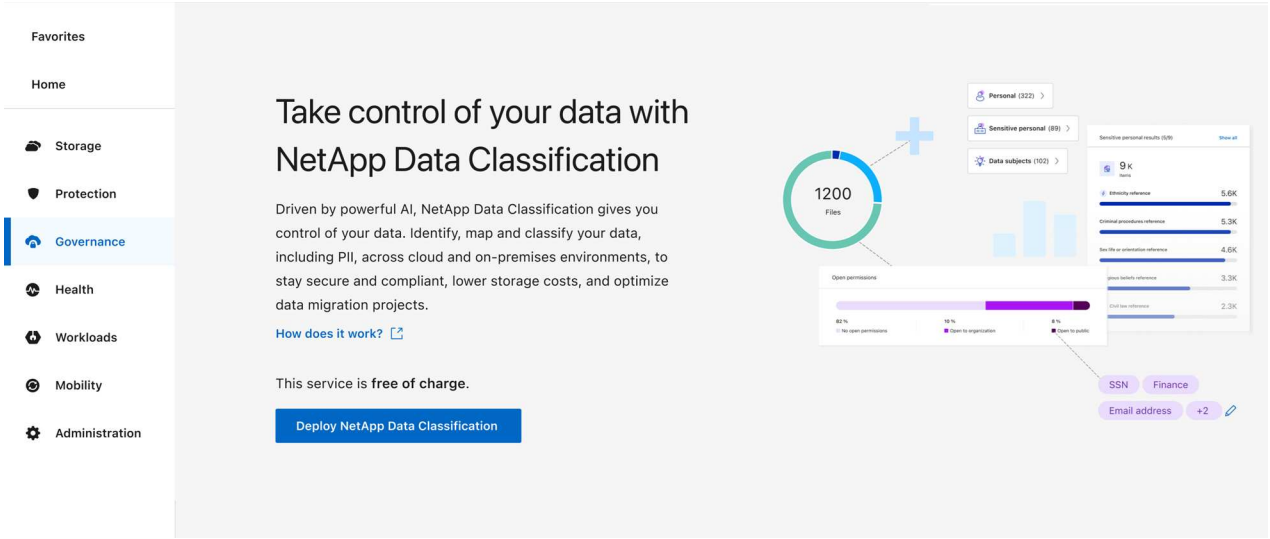


2. 選擇*部署*以啟動雲端部署精靈。
3. 精靈在執行部署步驟時會顯示進度。如果遇到任何問題，它將停止並提示輸入。
4. 當實例部署完畢並安裝資料分類後，選擇「繼續設定」進入「設定」頁面。

在 **Google Cloud** 部署

步驟

1. 從資料分類主頁中，選擇*治理>分類*。
2. 選擇*在本機或雲端部署分類*。



3. 選擇*部署*以啟動雲端部署精靈。
4. 精靈在執行部署步驟時會顯示進度。如果遇到任何問題，它將停止並提示輸入。
5. 當實例部署完畢並安裝資料分類後，選擇「繼續設定」進入「設定」頁面。

結果

控制台在您的雲端提供者中部署資料分類執行個體。

只要實例具有互聯網連接，控制台代理和資料分類軟體的升級就會自動進行。

下一步

您可以從設定頁面選擇要掃描的資料來源。

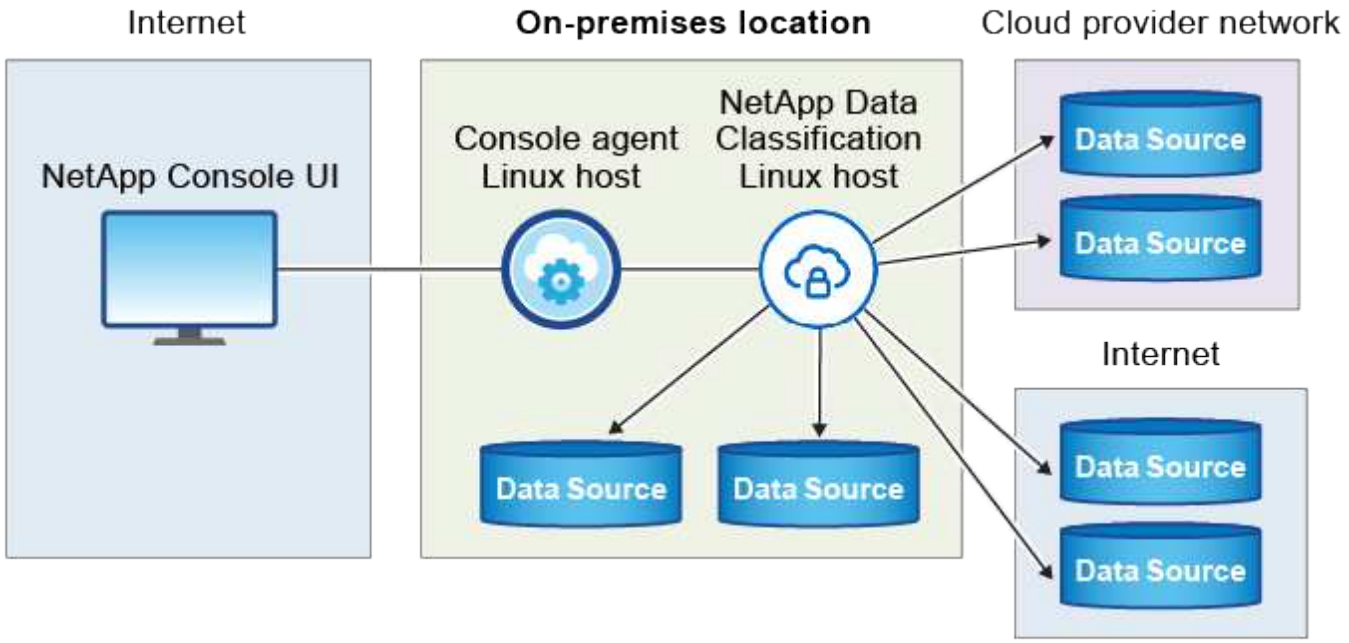
在可存取網際網路的主機上安裝**NetApp Data Classification**

要在您的網路中的 Linux 主機或具有 Internet 存取權限的雲端中的 Linux 主機上部署NetApp Data Classification，您需要在您的網路或雲端中手動部署 Linux 主機。

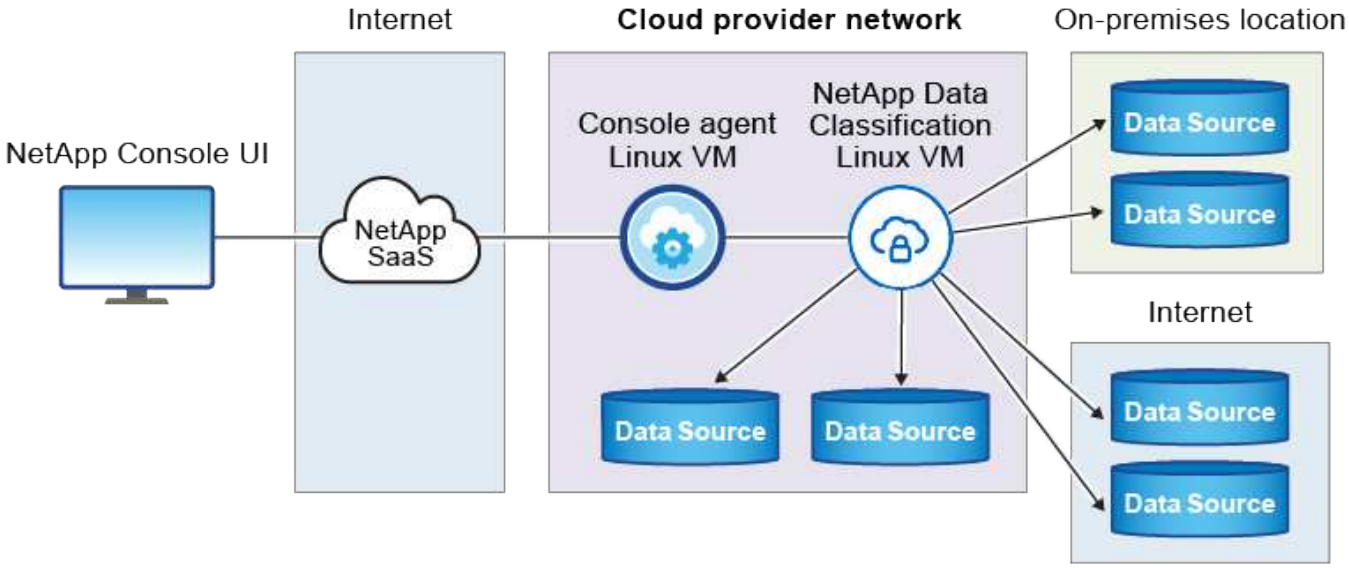
如果您希望使用同樣位於本機的資料分類實例來掃描本機ONTAP系統，則本機安裝是個不錯的選擇。這不是必需的。無論選擇哪一種安裝方法，軟體的功能都是相同的。

資料分類安裝腳本首先檢查系統和環境是否滿足所需的先決條件。如果所有先決條件都滿足，則安裝開始。如果您想獨立於執行資料分類安裝來驗證先決條件，您可以下載一個單獨的軟體包，該軟體包僅測試先決條件。"了解如何檢查您的 Linux 主機是否已準備好安裝資料分類" 。

您所在場所的 Linux 主機上的典型安裝具有以下元件和連接。



雲端 Linux 主機上的典型安裝具有以下元件和連接。



快速啟動

按照以下步驟快速開始，或向下捲動到其餘部分以獲取完整詳細資訊。

1 建立控制台代理

如果您還沒有控制台代理， "在本機部署控制台代理"在您的網路中的 Linux 主機上，或在雲端的 Linux 主機上。

您也可以與您的雲端提供者一起建立控制台代理程式。看 "在 AWS 中建立控制台代理"， "在 Azure 中建立控制台代理" ， 或者 "在 GCP 中建立控制台代理"。

2 審查先決條件

確保您的環境能夠滿足先決條件。這包括實例的出站互聯網存取、控制台代理和資料分類之間透過連接埠 443 的連接等等。查看完整列表 。

您還需要一個符合以下條件的 Linux 系統遵循要求。

3 下載並部署資料分類

從NetApp支援網站下載雲端資料分類軟體，並將安裝程式檔案複製到您打算使用的 Linux 主機。然後啟動安裝精靈並依照指示部署資料分類實例。

建立控制台代理

您需要先安裝控制台代理，然後才能安裝和使用資料分類。在大多數情況下，您可能會在嘗試啟動資料分類之前設定控制台代理，因為大多數 "控制台功能需要控制台代理"，但有些情況下您需要立即設定一個。

若要在您的雲端提供者環境中建立一個，請參閱 "在 AWS 中建立控制台代理"， "在 Azure 中建立控制台代理" ， 或者 "在 GCP 中建立控制台代理"。

在某些情況下，您必須使用部署在特定雲端提供者中的控制台代理：

- 在 AWS 或Amazon FSx for ONTAP中的Cloud Volumes ONTAP中掃描資料時，您可以使用 AWS 中的控制台代理程式。
- 在 Azure 中的Cloud Volumes ONTAP或Azure NetApp Files中掃描資料時，您可以使用 Azure 中的控制台代理程式。

  對於Azure NetApp Files，它必須部署在與您要掃描的磁碟區相同的區域中。

- 在 GCP 中的Cloud Volumes ONTAP中掃描資料時，您可以使用 GCP 中的控制台代理程式。

可以使用任何這些雲端控制台代理程式來掃描本機ONTAP系統、 NetApp檔案共用和資料庫帳戶。

請注意，您還可以 "在本機部署控制台代理"在您的網路中的 Linux 主機上或雲端中的 Linux 主機上。一些計劃在本機安裝資料分類的使用者可能也會選擇在本機安裝控制台代理程式。

安裝資料分類時，您將需要控制台代理系統的 IP 位址或主機名稱。如果您在您的場所安裝了控制台代理，您將獲得此資訊。如果控制台代理程式部署在雲端中，您可以從控制台中找到此資訊：選擇幫助圖標，然後選擇*支援*，然後選擇控制台代理。

準備 **Linux** 主機系統

資料分類軟體必須在滿足特定作業系統需求、RAM 需求、軟體需求等的主機上運作。 Linux 主機可以在您的網路中，也可以在雲端。

確保您可以保持資料分類運行。資料分類機器需要保持開啟以持續掃描您的資料。

- 資料分類必須運行在專用主機上。主機不能與其他應用程式或第三方軟體（例如防毒軟體）共用。
- 選擇與您打算使用資料分類掃描的資料集相符的大小。

系統大小	中央處理器	**RAM**（必須停用交換記憶體）	磁碟
超大	32 個 CPU	128 GB 內存	• / 上 1 TiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 上可用 895 GiB • /tmp 上 5 GiB • 對於 **Podman**，**/var/tmp** 上有 **30 GB**
大的	16 個 CPU	64 GB 內存	• / 上 500 GiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 或 Podman /var/lib/containers 上可用 400 GiB • /tmp 上 5 GiB • 對於 **Podman**，**/var/tmp** 上有 **30 GB**

- 在雲端為資料分類安裝部署運算執行個體時，建議您使用符合上述「大型」系統需求的系統：
  - **Amazon Elastic Compute Cloud (Amazon EC2)** 執行個體類型：「m6i.4xlarge」。"查看其他 AWS 執行個體類型"。
  - **Azure VM** 大小：「Standard_D16s_v3」。"查看其他 Azure 執行個體類型"。
  - **GCP** 機器類型：「n2-standard-16」。"查看其他 GCP 實例類型"。
- **UNIX** 資料夾權限：需要以下最低 UNIX 權限：

資料夾	最低權限
/tmp	`rwxrwxrwt`
/選擇	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/系統	`rwxr-xr-x`

- 作業系統:
  - 以下作業系統需要使用 Docker 容器引擎：
    - Red Hat Enterprise Linux 版本 7.8 與 7.9
    - Ubuntu 22.04（需要資料分類版本 1.23 或更高版本）
    - Ubuntu 24.04（需要資料分類版本 1.23 或更高版本）
  - 以下作業系統需要使用 Podman 容器引擎，並且需要資料分類版本 1.30 或更高版本：
    - Red Hat Enterprise Linux 版本 8.8、8.10、9.0、9.1、9.2、9.3、9.4、9.5 和 9.6。
  - 必須在主機系統上啟用進階向量擴充 (AVX2)。

- **Red Hat** 訂閱管理：主機必須在 Red Hat 訂閱管理中註冊。如果未註冊，系統將無法存取儲存庫來在安裝期間更新所需的第三方軟體。
- 附加軟體：安裝資料分類前，必須在主機上安裝以下軟體：
  - 根據您使用的作業系統，您需要安裝其中一個容器引擎：
    - Docker Engine 版本 19.3.1 或更高版本。 "查看安裝說明" 。
    - Podman 版本 4 或更高版本。若要安裝 Podman，請輸入(`sudo yum install podman netavark -y`) 。
- Python 版本 3.6 或更高版本。 "查看安裝說明" 。
  - **NTP** 注意事項： NetApp建議設定資料分類系統以使用網路時間協定 (NTP) 服務。資料分類系統和控制台代理系統之間的時間必須同步。
- **Firewalld** 注意事項：如果您打算使用 `firewalld`，我們建議您在安裝資料分類之前啟用它。運行以下命令進行配置 `firewalld`以便與資料分類相容：

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

如果您打算使用其他資料分類主機作為掃描器節點，請在此時將這些規則新增至您的主系統：

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

請注意，每次啟用或更新時都必須重新啟動 Docker 或 Podman `firewalld`設定.

安裝後，資料分類主機系統的 IP 位址無法變更。

啟用資料分類的出站互聯網訪問

資料分類需要出站網路存取。如果您的虛擬或實體網路使用代理伺服器進行網際網路訪問，請確保資料分類執行個體具有出站網際網路存取權限以聯絡下列端點。

端點	目的
\ https://api.console.netapp.com	與控制台的通信，其中包括NetApp帳戶。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	與控制台網站通信，實現集中用戶身份驗證。

端點	目的
\ https://support.compliance.api.bluexp.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https://dseasb33srnrn.cloudfront.net/ https://production.cloudflare.docker.com/.	提供對軟體映像、清單、範本的存取以及發送日誌和指標。
https://support.compliance.api.bluexp.netapp.com/	使NetApp能夠從稽核記錄中串流資料。
https://github.com/docker https://download.docker.com	提供docker安裝的必備包。
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	提供 Ubuntu 安裝的必備軟體包。

驗證所有必要的連接埠均已啟用

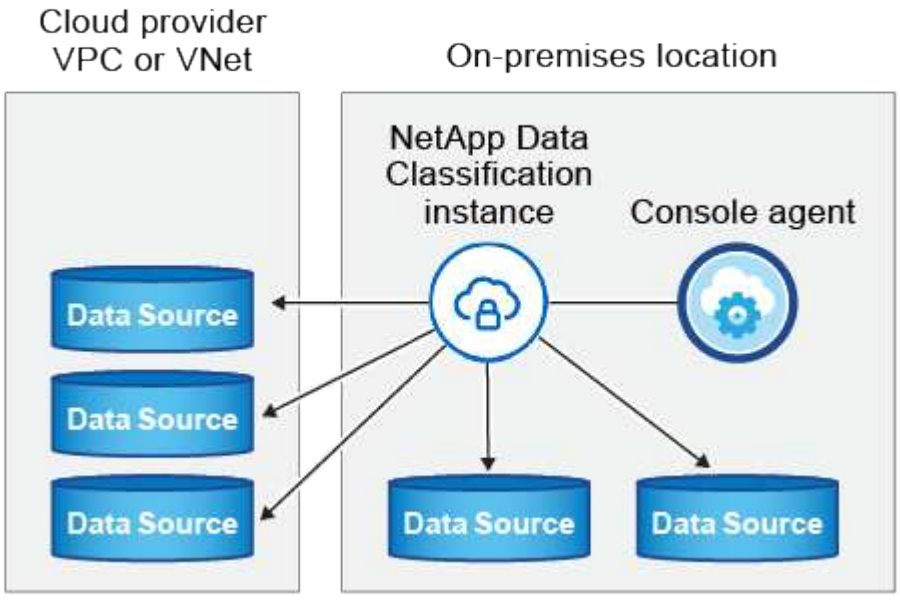
您必須確保所有必要的連接埠都已打開，以便在控制台代理程式、資料分類、Active Directory 和資料來源之間進行通訊。

連接類型	連接埠	描述
控制台代理<>資料分類	8080 (TCP)、443 (TCP) 和 80。9000	控制台代理程式的防火牆或路由規則必須允許透過連接埠 443 進出資料分類實例的入站和出站流量。確保連接埠 8080 已打開，以便您可以在控制台中看到安裝進度。如果 Linux 主機上使用防火牆，則 Ubuntu 伺服器內的內部進程需要連接埠 9000。
控制台代理<> ONTAP叢集 (NAS)	443（TCP）	控制台使用 HTTPS 發現ONTAP叢集。如果您使用自訂防火牆策略，則它們必須符合以下要求： • 控制台代理主機必須允許透過連接埠 443 進行出站 HTTPS 存取。如果控制台代理程式位於雲端中，則預先定義的防火牆或路由規則允許所有出站通訊。 • ONTAP叢集必須允許透過連接埠 443 進行入站 HTTPS 存取。預設的「mgmt」防火牆策略允許來自所有 IP 位址的入站 HTTPS 存取。如果您修改了此預設策略，或建立了自己的防火牆策略，則必須將 HTTPS 協定與該原則關聯並啟用從控制台代理主機的存取。

連接類型	連接埠	描述
資料分類 <> ONTAP集群	• 對於 NFS - 111 (TCP\UDP) 和 2049 (TCP\UDP) • 對於 CIFS - 139 (TCP\UDP) 和 445 (TCP\UDP)	資料分類需要與每個Cloud Volumes ONTAP子網路或本地ONTAP系統建立網路連線。 Cloud Volumes ONTAP的防火牆或路由規則必須允許來自資料分類實例的入站連線。 確保這些連接埠對資料分類實例開放： • 對於 NFS - 111 和 2049 • 對於 CIFS - 139 和 445 NFS 磁碟區匯出策略必須允許從資料分類實例進行存取。
資料分類<> Active Directory	389（TCP 和 UDP ）、636（TCP）、3268（TCP）和 3269（TCP）	您必須已經為公司使用者設定了 Active Directory。此外，資料分類需要 Active Directory 憑證來掃描 CIFS 磁碟區。 您必須具有 Active Directory 的資訊： • DNS 伺服器 IP 位址，或多個 IP 位址 • 伺服器的使用者名稱和密碼 • 網域名稱（Active Directory 名稱） • 您是否使用安全 LDAP (LDAPS) • LDAP 伺服器連接埠（LDAP 通常為 389，安全 LDAP 通常為 636）

在 **Linux** 主機上安裝資料分類

對於典型配置，您將在單一主機系統上安裝該軟體。請參閱此處的步驟 。



看準備 Linux 主機系統和審查先決條件了解部署資料分類之前的完整要求清單。

只要實例具有互聯網連接，資料分類軟體的升級就會自動進行。

當軟體安裝在本機時，資料分類目前無法掃描 S3 儲存桶、 Azure NetApp Files或 FSx for ONTAP 。在這些情況下，您需要在雲端中部署單獨的控制台代理程式和資料分類實例，並且 "在連接器之間切換"適用於不同的資料來源。

典型配置的單主機安裝

在單一本機上安裝資料分類軟體時，請查看要求並遵循下列步驟。

"觀看此視頻"了解如何安裝資料分類。

請注意，安裝資料分類時會記錄所有安裝活動。如果您在安裝過程中遇到任何問題，您可以查看安裝審計日誌的內容。它被寫給 `/opt/netapp/install_logs/`。

開始之前

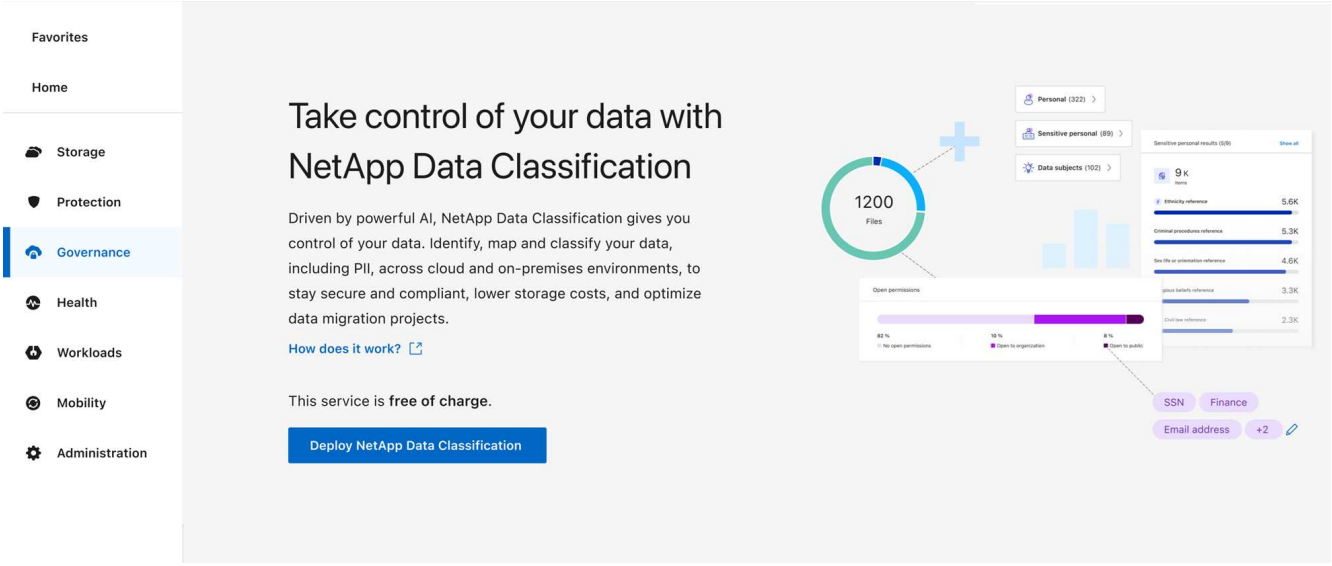
- 驗證您的 Linux 系統是否滿足主機需求。
- 驗證系統是否安裝了兩個必備軟體套件（Docker Engine 或 Podman 和 Python 3）。
- 確保您在 Linux 系統上擁有 root 權限。
- 如果您使用代理程式存取互聯網：
  - 您將需要代理伺服器資訊（IP 位址或主機名稱、連接連接埠、連接方案：https 或 http、使用者名稱和密碼）。
  - 如果代理程式正在執行 TLS 攔截，您需要知道資料分類 Linux 系統上儲存 TLS CA 憑證的路徑。
  - 代理必須是非透明的。資料分類目前不支援透明代理。
  - 該用戶必須是本機用戶。不支援網域用戶。
- 驗證您的離線環境是否符合要求權限和連線性。

步驟

1. 從下載資料分類軟體 "NetApp支援站點"。您應該選擇的檔案名稱為 **DATASENSE-INSTALLER-<version>.tar.gz**。
2. 將安裝程式檔案複製到您打算使用的 Linux 主機（使用 `scp`或其他方法）。
3. 在主機上解壓縮安裝程式文件，例如：

```
tar -xzf DATASENSE-INSTALLER-V1.25.0.tar.gz
```

4. 在控制台中，選擇*治理>分類*。
5. 選擇*在本機或雲端部署分類*。



6. 根據您是在雲端中準備的實例上還是在本機準備的實例上安裝資料分類，選擇適當的*部署*選項來啟動資料分類安裝。
7. 將顯示「在本機部署資料分類」對話方塊。複製提供的命令（例如： `sudo ./install.sh -a 12345 -c 27AG75 -t 2198qq` ）並將其貼到文字檔案中，以便稍後使用。然後選擇*關閉*以關閉對話框。
8. 在主機上，輸入您複製的命令，然後按照一系列提示進行操作，或者您可以提供包含所有必需參數的完整命令作為命令列參數。

   請注意，安裝程式會執行預檢查以確保您的系統和網路要求滿足，以便成功安裝。 "觀看此視頻"了解預檢資訊和意義。

根據提示輸入參數：	輸入完整命令：
a. 貼上從步驟 7 複製的命令： `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token>` 如果您在雲端實例上安裝（而不是在您的本地），請新增 `--manual-cloud-install <cloud_provider>`。 b. 輸入資料分類主機的 IP 位址或主機名，以便控制台代理系統可以存取它。 c. 輸入控制台代理主機的 IP 位址或主機名，以便資料分類系統可以存取它。 d. 根據提示輸入代理詳細資料。如果您的控制台代理已經使用代理，則無需在此處再次輸入此信息，因為資料分類將自動使用控制台代理所使用的代理。	或者，您可以提前建立整個命令，提供必要的主機和代理參數： `sudo ./install.sh -a <account_id> -c <client_id> -t <user_token> --host <ds_host> --manager-host <cm_host> --manual-cloud-install <cloud_provider> --proxy-host <proxy_host> --proxy-port <proxy_port> --proxy-scheme <proxy_scheme> --proxy-user <proxy_user> --proxy-password <proxy_password> --cacert-folder-path <ca_cert_dir>`

變數值：

- *account_id* = NetApp帳號 ID

- *client_id* = 控制台代理客戶端 ID（如果客戶端 ID 中沒有後綴“clients”，則新增後綴）
- *user_token* = JWT 使用者存取權令牌
- *ds_host* = 資料分類 Linux 系統的 IP 位址或主機名稱。
- *cm_host* = 控制台代理系統的 IP 位址或主機名稱。
- *cloud_provider* = 在雲端實例上安裝時，根據雲端提供者輸入「AWS」、「Azure」或「Gcp」。
- *proxy_host* = 如果主機位於代理伺服器後面，則為代理伺服器的 IP 或主機名稱。
- *proxy_port* = 連接到代理伺服器的連接埠（預設為 80）。
- *proxy_scheme* = 連接方案：https 或 http（預設 http）。
- *proxy_user* = 如果需要基本驗證，則經過驗證的使用者連接到代理伺服器。使用者必須是本機使用者 - 不支援網域使用者。
- *proxy_password* = 您指定的使用者名稱的密碼。
- *ca_cert_dir* = 資料分類 Linux 系統上包含附加 TLS CA 憑證包的路徑。僅當代理執行 TLS 攔截時才需要。

結果

資料分類安裝程式安裝套件、註冊安裝並安裝資料分類。安裝可能需要 10 到 20 分鐘。

如果主機和控制台代理執行個體之間透過連接埠 8080 建立連接，您將在控制台的「資料分類」標籤中看到安裝進度。

下一步

您可以從設定頁面選擇要掃描的資料來源。

在沒有網路存取的 **Linux** 主機上安裝**NetApp Data Classification**

在沒有網路存取權限的本機站點的 Linux 主機上安裝NetApp Data Classification稱為_私有模式_。這種安裝類型使用安裝腳本，與NetApp Console SaaS 層沒有連線。

BlueXP私有模式（傳統BlueXP介面）通常用於沒有網路連線的本機環境和安全雲端區域，其中包括 AWS Secret Cloud、AWS Top Secret Cloud 和 Azure IL6。NetApp繼續透過傳統的BlueXP介面支援這些環境。有關舊版BlueXP介面中的私有模式文檔，請參閱"BlueXP私人模式的 PDF 文檔"。

檢查您的 **Linux** 主機是否已準備好安裝**NetApp Data Classification**

在 Linux 主機上手動安裝NetApp Data Classification之前，可以選擇在主機上執行腳本以驗證安裝資料分類的所有先決條件是否都已具備。您可以在網路中的 Linux 主機上或雲端中的 Linux 主機上執行此腳本。主機可以連接到互聯網，或者主機可以駐留在沒有互聯網訪問的站點（暗站）。

資料分類安裝腳本包含一個測試腳本，以確保您的環境符合要求。您可以單獨執行此腳本來驗證 Linux 主機是否已準備就緒，然後再執行安裝腳本。

入門

您將執行以下任務。

- 如果您尚未安裝控制台代理，則可以選擇安裝。您可以在未安裝控制台代理的情況下執行測試腳本，但腳本會檢查控制台代理程式和資料分類主機之間的連線 - 因此建議您安裝控制台代理程式。
- 準備主機並驗證其是否滿足所有要求。
- 啟用資料分類主機的出站網際網路存取。
- 驗證所有系統上的所有必要連接埠是否都已啟用。
- 下載並執行先決條件測試腳本。

建立控制台代理

您需要先安裝控制台代理，然後才能安裝和使用資料分類。但是，您可以在沒有控制台代理的情況下執行先決條件腳本。

你可以 "在本機安裝控制台代理" 在您網路中的 Linux 主機上，或是在雲端的 Linux 主機上。如果控制台代理程式安裝在本機，您也可以在本機安裝資料分類。

若要在您的雲端提供者環境中建立控制台代理，請參閱：

- "在 AWS 中建立控制台代理"
- "在 Azure 中建立控制台代理"
- "在 GCP 中建立控制台代理"

執行先決條件腳本時，需要控制台代理系統的 IP 位址或主機名稱。如果您在本機安裝了控制台代理，則可以取得此資訊。如果控制台代理程式部署在雲端，您可以從控制台中找到此資訊：選擇說明圖標，然後選擇「支援」；在「代理程式和審核」部分，選擇「前往代理程式」。

驗證主機需求

資料分類軟體必須運行在滿足特定作業系統要求、記憶體需求和軟體要求的主機上。

- 資料分類必須運行在專用主機上。主機不能與其他應用程式或第三方軟體（例如防毒軟體）共用。
- 選擇與您打算使用資料分類掃描的資料集相符的大小。

系統大小	中央處理器	**RAM**（必須停用交換記憶體）	磁碟
超大	32 個 CPU	128 GB 內存	• / 上 1 TiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 上可用 895 GiB • /tmp 上 5 GiB • 對於 **Podman**，**/var/tmp** 上有 **30 GB**

系統大小	中央處理器	**RAM**（必須停用交換記憶體）	磁碟
大的	16 個 CPU	64 GB 內存	• / 上 500 GiB SSD，或 /opt 上 100 GiB 可用 • /var/lib/docker 或 Podman /var/lib/containers 上可用 400 GiB • /tmp 上 5 GiB • 對於 **Podman**，**/var/tmp** 上有 **30 GB**

- 在雲端為資料分類安裝部署運算執行個體時，建議您使用符合上述「大型」系統需求的系統：
  - **Amazon Elastic Compute Cloud (Amazon EC2)** 執行個體類型：「m6i.4xlarge」。"查看其他 AWS 執行個體類型"。
  - **Azure VM** 大小：「Standard_D16s_v3」。"查看其他 Azure 執行個體類型"。
  - **GCP** 機器類型：「n2-standard-16」。"查看其他 GCP 實例類型"。
- **UNIX** 資料夾權限：需要以下最低 UNIX 權限：

資料夾	最低權限
/tmp	`rwxrwxrwt`
/選擇	`rwxr-xr-x`
/var/lib/docker	`rwx------`
/usr/lib/systemd/系統	`rwxr-xr-x`

- 作業系統:
  - 以下作業系統需要使用 Docker 容器引擎：
    - Red Hat Enterprise Linux 版本 7.8 與 7.9
    - Ubuntu 22.04（需要資料分類版本 1.23 或更高版本）
    - Ubuntu 24.04（需要資料分類版本 1.23 或更高版本）
  - 以下作業系統需要使用 Podman 容器引擎，並且需要資料分類版本 1.30 或更高版本：
    - Red Hat Enterprise Linux 版本 8.8、8.10、9.0、9.1、9.2、9.3、9.4、9.5 和 9.6。
  - 必須在主機系統上啟用進階向量擴充 (AVX2)。
- **Red Hat** 訂閱管理：主機必須在 Red Hat 訂閱管理中註冊。如果未註冊，系統將無法存取儲存庫來在安裝期間更新所需的第三方軟體。
- 附加軟體：安裝資料分類前，必須在主機上安裝以下軟體：
  - 根據您使用的作業系統，您需要安裝其中一個容器引擎：
    - Docker Engine 版本 19.3.1 或更高版本。"查看安裝說明"。
    - Podman 版本 4 或更高版本。若要安裝 Podman，請輸入(`sudo yum install podman netavark -y`)。

- Python 版本 3.6 或更高版本。 "查看安裝說明" 。
  - **NTP** 注意事項： NetApp建議設定資料分類系統以使用網路時間協定 (NTP) 服務。資料分類系統和控制台代理系統之間的時間必須同步。
- **Firewalld** 注意事項：如果您打算使用 `firewalld`，我們建議您在安裝資料分類之前啟用它。運行以下命令進行配置 `firewalld`以便與資料分類相容：

```
firewall-cmd --permanent --add-service=http
firewall-cmd --permanent --add-service=https
firewall-cmd --permanent --add-port=80/tcp
firewall-cmd --permanent --add-port=8080/tcp
firewall-cmd --permanent --add-port=443/tcp
firewall-cmd --reload
```

如果您打算使用其他資料分類主機作為掃描器節點（在分散式模型中），請在此時將這些規則新增至您的主系統：

```
firewall-cmd --permanent --add-port=2377/tcp
firewall-cmd --permanent --add-port=7946/udp
firewall-cmd --permanent --add-port=7946/tcp
firewall-cmd --permanent --add-port=4789/udp
```

請注意，每次啟用或更新時都必須重新啟動 Docker 或 Podman `firewalld`設定.

啟用資料分類的出站互聯網訪問

資料分類需要出站網路存取。如果您的虛擬或實體網路使用代理伺服器進行網際網路訪問，請確保資料分類執行個體具有出站網際網路存取權限以聯絡下列端點。

對於安裝在沒有網路連線的站點的主機系統，則不需要本節。

端點	目的
\ https://api.console.netapp.com	與控制台服務（包括NetApp帳戶）的通訊。
\ https://netapp-cloud-account.auth0.com \ https://auth0.com	與控制台網站通信，實現集中用戶身份驗證。
\ https://support.compliance.api.console.netapp.com/ \ https://hub.docker.com \ https://auth.docker.io \ https://registry-1.docker.io \ https://index.docker.io/ \ https https://dseasb33srnrn.cloudfront.net/ // https://production.cloudflare.docker.com/	提供對軟體映像、清單、範本的存取以及發送日誌和指標。
\ https://support.compliance.api.console.netapp.com/	使NetApp能夠從稽核記錄中串流資料。

端點	目的
https://github.com/docker https://download.docker.com	提供docker安裝的必備包。
\ http://packages.ubuntu.com/ \ http://archive.ubuntu.com	提供 Ubuntu 安裝的必備軟體包。

驗證所有必要的連接埠均已啟用

您必須確保所有必要的連接埠都已打開，以便在控制台代理程式、資料分類、Active Directory 和資料來源之間進行通訊。

連接類型	連接埠	描述
控制台代理<>資料分類	8080 (TCP)、443 (TCP)和 80。9000	控制台代理程式的防火牆或路由規則必須允許透過連接埠 443 進出資料分類實例的入站和出站流量。確保連接埠 8080 已打開，以便您可以在控制台中看到安裝進度。如果 Linux 主機上使用防火牆，則 Ubuntu 伺服器內的內部進程需要連接埠 9000。
控制台代理<> ONTAP叢集 (NAS)	443（TCP）	控制台使用 HTTPS 發現ONTAP叢集。如果您使用自訂防火牆策略，控制台代理主機必須允許透過連接埠 443 進行出站 HTTPS 存取。如果控制台代理程式位於雲端中，則預先定義的防火牆或路由規則允許所有出站通訊。

執行資料分類先決條件腳本

請依照以下步驟執行資料分類先決條件腳本。

"觀看此視頻"了解如何執行先決條件腳本並解釋結果。

開始之前

- 驗證您的 Linux 系統是否滿足主機需求。
- 驗證系統是否安裝了兩個必備軟體套件（Docker Engine 或 Podman 和 Python 3）。
- 確保您在 Linux 系統上擁有 root 權限。

步驟

1. 從下載資料分類先決條件腳本 "NetApp支援站點"。您應該選擇的檔案名稱為 **standalone-pre-requisite-tester-<version>**。
2. 將檔案複製到您計劃使用的 Linux 主機（使用 `scp`或其他方法）。
3. 分配運行腳本的權限。

```
chmod +x standalone-pre-requisite-tester-v1.25.0
```

4. 使用以下命令運行腳本。

```
./standalone-pre-requisite-tester-v1.25.0 <--darksite>
```

僅當您在沒有網際網路存取的主機上執行腳本時才新增選項「--darksite」。當主機未連接到網際網路時，某些先決條件測試將被跳過。

5. 腳本會提示您輸入資料分類主機的 IP 位址。
   - 輸入 IP 位址或主機名稱。
6. 該腳本會提示您是否安裝了控制台代理程式。
   - 如果您沒有安裝控制台代理，請輸入 **N**。
   - 如果您確實安裝了控制台代理，請輸入 **Y**。然後輸入控制台代理的 IP 位址或主機名，以便測試腳本可以測試此連線。
7. 該腳本在系統上執行各種測試，並在運行過程中顯示結果。完成後，它會將會話日誌寫入名為 `prerequisites-test-<timestamp>.log`在目錄中 `/opt/netapp/install_logs`。

結果

如果所有先決條件測試均成功運行，您可以在準備就緒後在主機上安裝資料分類。

如果發現任何問題，則將其歸類為“建議”或“需要”修復。建議的問題通常是會使資料分類掃描和分類任務運作速度變慢的項目。這些項目不需要糾正 - 但您可能需要解決它們。

如果您有任何「必要」問題，您應該修復這些問題並再次執行先決條件測試腳本。

啟動資料來源掃描

使用**NetApp Data Classification**掃描資料來源

NetApp Data Classification會掃描您選擇的儲存庫（磁碟區、資料庫模式或其他使用者資料）中的數據，以識別個人和敏感資料。然後，資料分類會對應您的組織資料、對每個文件進行分類並識別資料中的預定義模式。掃描結果是個人資訊、敏感個人資訊、資料類別和文件類型的索引。

初步掃描後，資料分類將以循環方式持續掃描您的資料以偵測增量變化。這就是為什麼保持實例運行很重要。

您可以在磁碟區層級或資料庫模式層級啟用和停用掃描。

映射掃描和分類掃描之間有什麼區別

您可以在資料分類中進行兩種類型的掃描：

- 僅映射掃描僅提供資料的進階概覽，並在選定的資料來源上執行。僅映射掃描比映射和分類掃描花費的時間更少，因為它們不存取文件來查看其中的資料。您可能希望首先執行此操作來確定研究領域，然後對這些領域執行地圖和分類掃描。
- 地圖和分類掃描 為您的資料提供深層掃描。

下表顯示了一些差異：

特徵	映射和分類掃描	僅映射掃描
掃描速度	慢的	快速地
定價	自由的	自由的
容量	限制為 500 TiB*	限制為 500 TiB*
文件類型和已使用容量列表	是的	是的
文件數量和已用容量	是的	是的
文件的年齡和大小	是的	是的
能夠運行"數據映射報告"	是的	是的
數據調查頁面查看文件詳細信息	是的	不
在文件中搜尋名稱	是的	不
創造"已儲存的查詢"提供自訂搜尋結果	是的	不
能夠運行其他報告	是的	不
能夠查看文件中的元資料**	不	是的

* 資料分類不會對其可掃描的資料量施加限制。每個控制台代理程式支援掃描和顯示 500 TiB 的資料。要掃描超過 500 TiB 的數據，"安裝另一個控制台代理"然後"部署另一個資料分類實例"。 + 控制台 UI 顯示來自單一連接器的資料。有關查看來自多個控制台代理的資料的提示，請參閱"使用多個控制台代理"。

** 在映射掃描期間從檔案中提取以下元資料：

- 系統
- 系統類型
- 儲存庫
- 文件類型
- 已用容量
- 文件數量
- 文件大小
- 文件創建
- 文件上次訪問
- 文件上次修改時間
- 文件發現時間
- 權限擷取

治理儀表板差異：

特徵	地圖和分類	地圖
過時的數據	是的	是的
非業務數據	是的	是的
重複文件	是的	是的
預定義已儲存的查詢	是的	不
預設儲存的查詢	是的	是的
DDA 報告	是的	是的
地圖報告	是的	是的
靈敏度等級檢測	是的	不
具有廣泛權限的敏感數據	是的	不
開放權限	是的	是的
數據時代	是的	是的
數據大小	是的	是的
類別	是的	不
文件類型	是的	是的

合規性儀表板差異：

特徵	地圖和分類	地圖
個人資訊	是的	不
敏感個人資訊	是的	不
隱私風險評估報告	是的	不
HIPAA 報告	是的	不
PCI DSS 報告	是的	不

調查過濾器差異：

特徵	地圖和分類	地圖
已儲存的查詢	是的	是的
系統類型	是的	是的
系統	是的	是的
儲存庫	是的	是的
文件類型	是的	是的
文件大小	是的	是的
創建時間	是的	是的
發現時間	是的	是的
上次修改時間	是的	是的
上次訪問	是的	是的
開放權限	是的	是的
檔案目錄路徑	是的	是的
類別	是的	不
敏感度等級	是的	不
識別符數量	是的	不
個人資料	是的	不
敏感個人數據	是的	不
資料主體	是的	不
重複項	是的	是的
分類狀態	是的	狀態始終為“見解有限”
掃描分析事件	是的	是的
文件哈希	是的	是的
有存取權限的使用者數	是的	是的
使用者/群組權限	是的	是的
文件所有者	是的	是的
目錄類型	是的	是的

使用**NetApp Data Classification**掃描**Amazon FSx for ONTAP**卷

完成幾個步驟即可使用NetApp Data Classification掃描Amazon FSx for ONTAP磁碟區。

開始之前

- 您需要 AWS 中一個活動的控制台代理來部署和管理資料分類。
- 建立系統時選擇的安全群組必須允許來自資料分類實例的流量。您可以使用連接到 FSx for ONTAP檔案系統的 ENI 找到關聯的安全性群組，並使用 AWS 管理主控台對其進行編輯。

  "Linux 執行個體的 AWS 安全群組"

  "Windows 執行個體的 AWS 安全性群組"

  "AWS 彈性網路介面 (ENI)"

- 確保以下連接埠對資料分類實例開放：
  - 對於 NFS – 連接埠 111 和 2049。
  - 對於 CIFS – 連接埠 139 和 445。

部署資料分類實例

"部署資料分類"如果尚未部署實例。

您應該在與 AWS 主控台代理程式和要掃描的 FSx 磁碟區相同的 AWS 網路中部署資料分類。

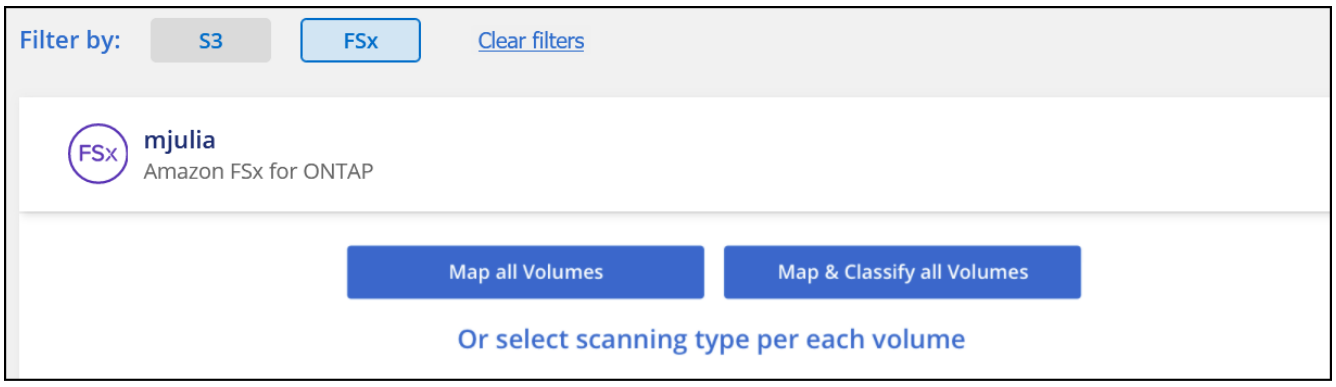
*注意：*掃描 FSx 磁碟區時，目前不支援在本機位置部署資料分類。

只要實例具有互聯網連接，資料分類軟體的升級就會自動進行。

在您的系統中啟用資料分類

您可以為 FSx for ONTAP磁碟區啟用資料分類。

1. 從NetApp Console，治理 > 分類。
2. 從資料分類選單中，選擇*配置*。



3. 選擇如何掃描每個系統中的磁碟區。"了解映射和分類掃描"：
   - 若要對應所有捲，請選擇*映射所有磁碟區*。
   - 若要對應和分類所有捲，請選擇*映射和分類所有捲*。
   - 若要自訂每個磁碟區的掃描，請選擇*或選擇每個磁碟區的掃描類型*，然後選擇要對應和/或分類的磁碟

區。

4. 在確認對話方塊中，選擇「核准」以使資料分類開始掃描您的磁碟區。

結果

資料分類開始掃描您在系統中選擇的磁碟區。一旦資料分類完成初始掃描，結果將在合規性儀表板中提供。所需時間取決於資料量——可能是幾分鐘或幾小時。您可以透過導航到配置選單然後選擇系統配置來追蹤初始掃描的進度。在進度條中追蹤每次掃描的進度；您可以將滑鼠懸停在進度條上，以查看相對於磁碟區中總檔案數的掃描檔案數。

- 預設情況下，如果資料分類在 CIFS 中沒有寫入屬性權限，或者在 NFS 中沒有寫入權限，系統將不會掃描磁碟區中的文件，因為資料分類無法將「上次存取時間」還原為原始時間戳記。如果您不介意上次造訪時間是否重設，請選擇*或為每個磁碟區選擇掃描類型*。結果頁面有一個您可以啟用的設置，以便資料分類可以掃描卷，而不管權限如何。
- 資料分類僅掃描卷下的一個檔案共用。如果您的磁碟區中有多個共享，則需要將這些其他共享作為共享群組單獨掃描。"查看有關此數據分類限制的更多詳細信息" 。

驗證資料分類是否有權存取卷

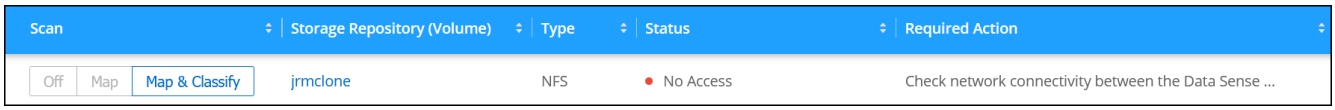
透過檢查網路、安全群組和匯出策略，確保資料分類可以存取磁碟區。

您需要向資料分類提供 CIFS 憑證，以便它可以存取 CIFS 磁碟區。

步驟

1. 從資料分類選單中，選擇*配置*。
2. 在配置頁面上，選擇*查看詳細資訊*以查看狀態並修正任何錯誤。

   例如，下圖顯示由於資料分類實例和磁碟區之間的網路連線問題，資料分類無法掃描磁碟區。



3. 確保資料分類實例與包含 FSx for ONTAP磁碟區的每個網路之間存在網路連線。

   對於 FSx for ONTAP，資料分類只能掃描與控制台位於相同區域的磁碟區。

4. 確保 NFS 磁碟區匯出策略包含資料分類執行個體的 IP 位址，以便它可以存取每個磁碟區上的資料。
5. 如果您使用 CIFS，請向資料分類提供 Active Directory 憑證，以便它可以掃描 CIFS 磁碟區。

   a. 從資料分類選單中，選擇*配置*。

   b. 對於每個系統，選擇*編輯 CIFS 憑證*並輸入資料分類存取系統上的 CIFS 磁碟區所需的使用者名稱和密碼。

      憑證可以是唯讀的，但提供管理員憑證可確保資料分類可以讀取任何需要提升權限的資料。憑證儲存在資料分類實例上。

      如果您想要確保檔案的「上次存取時間」不會因資料分類掃描而改變，建議使用者在 CIFS 中具有寫入屬性權限或在 NFS 中具有寫入權限。如果可能，請將 Active Directory 使用者設定為組織中具有所有檔案權限的父群組的一部分。

輸入憑證後，您應該會看到一條訊息，表示所有 CIFS 磁碟區均已成功驗證。

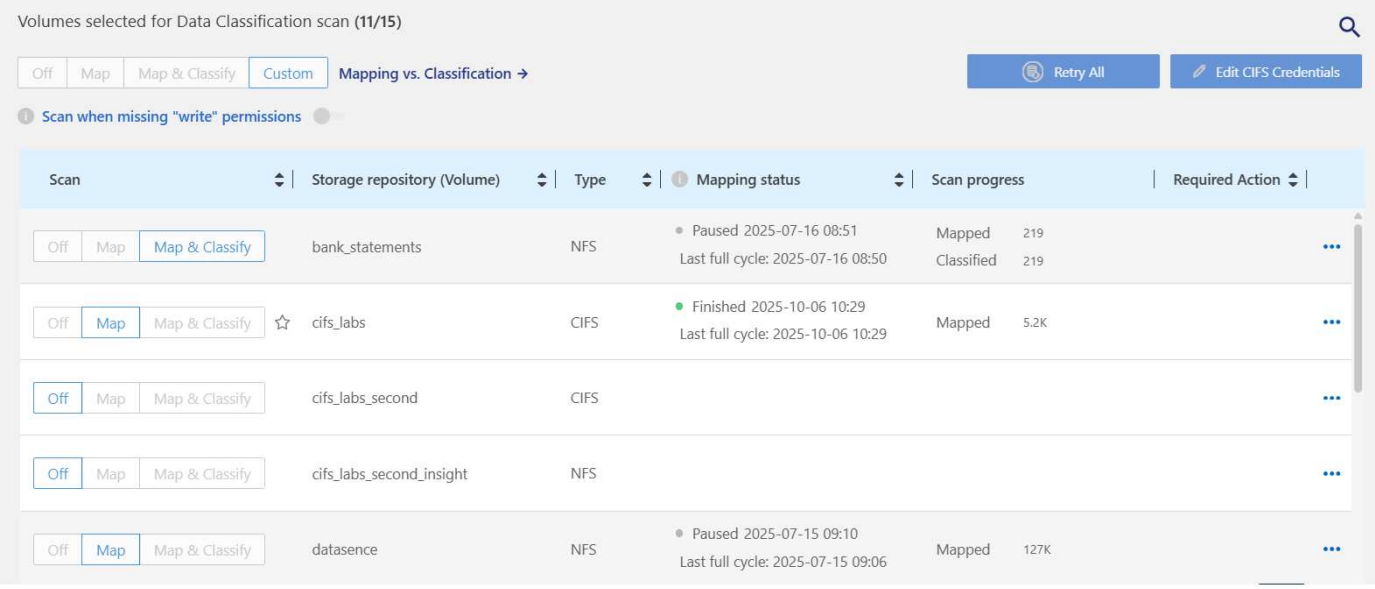
啟用和停用磁碟區上的掃描

您可以隨時從設定頁面啟動或停止任何系統上的掃描。您也可以將掃描從僅映射掃描切換到映射和分類掃描，反之亦然。建議您掃描系統中的所有磁碟區。

只有當您在標題區域中選擇了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增至系統的新磁碟區。當在標題區域設定為*自訂*或*關閉*時，您需要在系統中新增的每個新磁碟區上啟動對應和/或完整掃描。

頁面頂部的「缺少「寫入」權限時掃描」開關預設為停用狀態。這意味著，如果資料分類在 CIFS 中沒有寫入屬性權限或在 NFS 中沒有寫入權限，系統將不會掃描文件，因為資料分類無法將「上次存取時間」恢復為原始時間戳記。如果您不介意是否重設上次造訪時間，請開啟開關，無論權限為何，都會掃描所有檔案。"了解更多"。

只有當您在標題區域中設定了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增到系統的新磁碟區。當所有磁碟區的設定都是「自訂」或「關閉」時，您需要為新增的每個新磁碟區手動啟動掃描。



步驟

1. 從資料分類選單中，選擇*配置*。
2. 選擇一個系統，然後選擇*配置*。
3. 若要啟用或停用所有磁碟區的掃描，請在所有磁碟區上方的標題中選擇對應、對應和分類或關閉。

   若要啟用或停用對單一卷的掃描，請在清單中找到該卷，然後選擇卷名稱旁的映射、映射和分類或關閉。

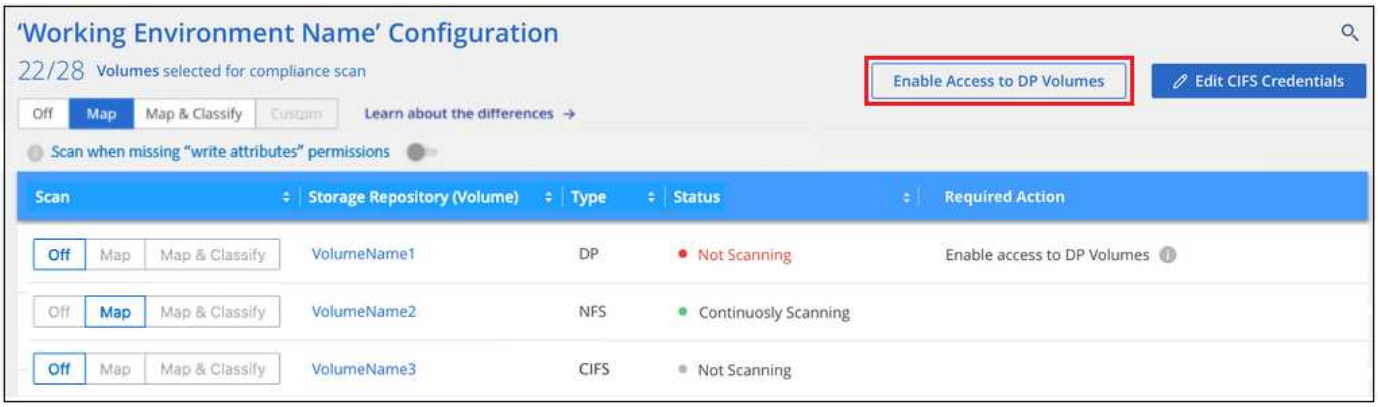
結果

當您啟用掃描時，資料分類將開始掃描您在系統中選擇的磁碟區。一旦資料分類開始掃描，結果就會開始出現在合規性儀表板中。掃描完成時間取決於資料量，從幾分鐘到幾小時不等。

掃描資料保護卷

預設情況下，不會掃描資料保護 (DP) 卷，因為它們未暴露在外部，且資料分類無法存取它們。這些是來自 FSx for ONTAP檔案系統的SnapMirror操作的目標磁碟區。

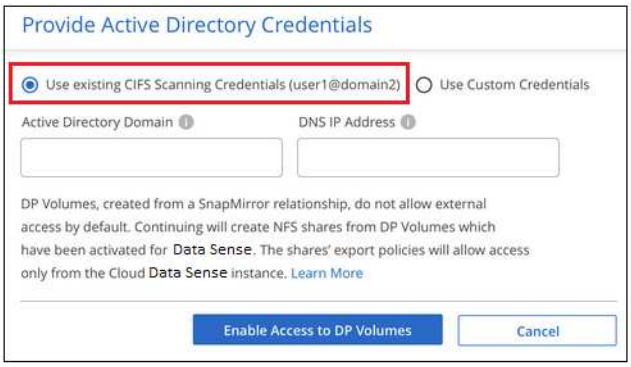
最初，磁碟區清單將這些磁碟區標識為_類型_ **DP**，其_狀態_ 未掃描*和_所需操作_ *啟用對 **DP** 磁碟區的存取。

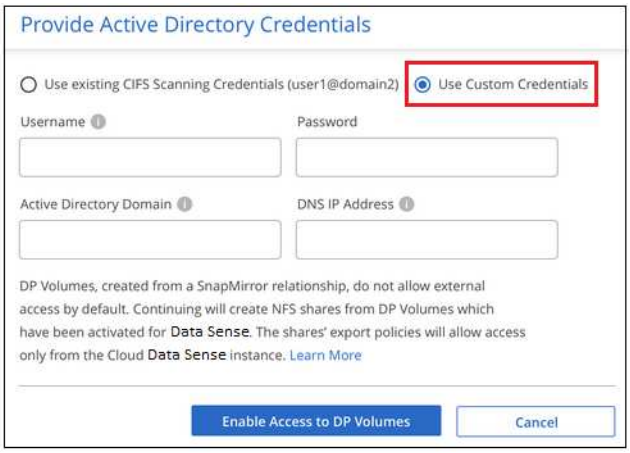


步驟

如果要掃描這些資料保護磁碟區：

1. 從資料分類選單中，選擇*配置*。
2. 選擇頁面頂部的「啟用對 DP 磁碟區的存取」*。
3. 查看確認訊息並再次選擇*啟用對 DP 磁碟區的存取*。
   - 最初在來源 FSx for ONTAP檔案系統中建立為 NFS 磁碟區的磁碟區已啟用。
   - 最初在來源 FSx for ONTAP檔案系統中建立為 CIFS 磁碟區的磁碟區會要求您輸入 CIFS 憑證來掃描這些 DP 磁碟區。如果您已經輸入了 Active Directory 憑證，以便資料分類可以掃描 CIFS 卷，您可以使用這些憑證，或者您可以指定一組不同的管理員憑證。





4. 啟動您想要掃描的每個 DP 磁碟區。

結果

一旦啟用，資料分類將從每個啟動掃描的 DP 磁碟區建立一個 NFS 共用。共享導出策略僅允許從資料分類實例進行存取。

如果您在最初啟用對 DP 磁碟區的存取權時沒有 CIFS 資料保護卷，後來又添加了一些，則按鈕 啟用對 **CIFS DP** 的存取 將出現在設定頁面的頂部。選擇此按鈕並新增 CIFS 憑證以啟用對這些 CIFS DP 磁碟區的存取。

Active Directory 憑證僅在第一個 CIFS DP 磁碟區的儲存 VM 中註冊，因此該 SVM 上的所有 DP 磁碟區都將掃描。駐留在其他 SVM 上的任何磁碟區都不會註冊 Active Directory 憑證，因此不會掃描這些 DP 磁碟區。

使用NetApp Data Classification掃描Azure NetApp Files卷

完成幾個步驟即可開始使用適用於Azure NetApp Files的NetApp Data Classification。

發現要掃描的**Azure NetApp** 檔案系統**Discover the Azure NetApp Files system that you want to scan**

如果要掃描的Azure NetApp Files系統尚未作為系統出現在NetApp Console中，"將其新增至系統頁面" 。

部署資料分類實例

"部署資料分類"如果尚未部署實例。

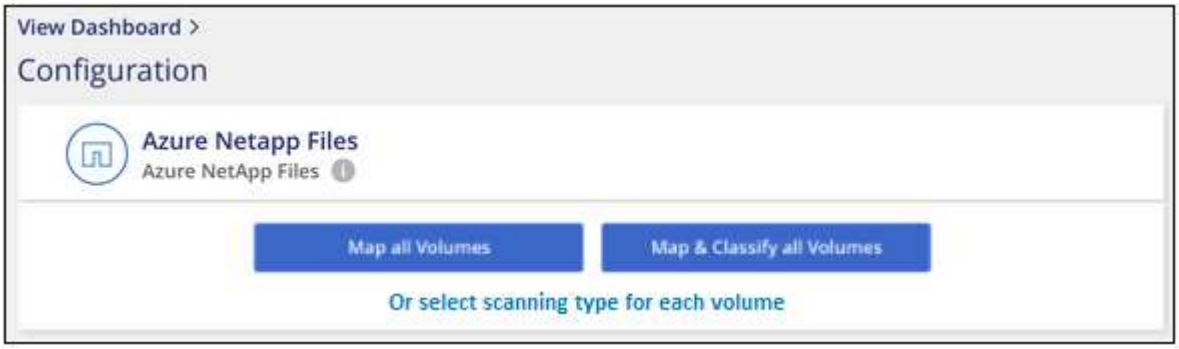
掃描Azure NetApp Files區時，資料分類必須部署在雲端中，並且必須部署在與要掃描的磁碟區相同的區域中。

*注意：*掃描Azure NetApp Files區時，目前不支援在本機位置部署資料分類。

在您的系統中啟用資料分類

您可以在Azure NetApp Files區上啟用資料分類。

1. 從資料分類選單中，選擇*配置*。



2. 選擇如何掃描每個系統中的磁碟區。"了解映射和分類掃描"：
   - 若要對應所有捲，請選擇*映射所有磁碟區*。
   - 若要對應和分類所有捲，請選擇*映射和分類所有捲*。
   - 若要自訂每個磁碟區的掃描，請選擇*或選擇每個磁碟區的掃描類型*，然後選擇要對應或對應和分類的磁碟區。

     看啟用或停用磁碟區掃描了解詳情。

3. 在確認對話方塊中，選擇*批准*。

結果

資料分類開始掃描您在系統中選擇的磁碟區。一旦資料分類完成初始掃描，結果就會顯示在合規性儀表板中。所需時間取決於資料量——可能是幾分鐘或幾小時。您可以透過導航到配置選單然後選擇系統配置來追蹤初始掃描的進度。資料分類顯示每次掃描的進度條。您可以將滑鼠懸停在進度條上，以查看相對於磁碟區中檔案總數的已掃描檔案數。

- 預設情況下，如果資料分類在 CIFS 中沒有寫入屬性權限，或者在 NFS 中沒有寫入權限，系統將不會掃描磁碟區中的文件，因為資料分類無法將「上次存取時間」還原為原始時間戳記。如果您不介意上次造訪時間是否重設，請選擇*或為每個磁碟區選擇掃描類型*。結果頁面有一個您可以啟用的設置，以便資料分類可以掃描卷，而不管權限如何。
- 資料分類僅掃描卷下的一個檔案共用。如果您的磁碟區中有多個共享，則需要將這些其他共享作為共享群組單獨掃描。"了解此資料分類限制" 。

驗證資料分類是否有權存取卷

透過檢查網路、安全群組和匯出策略，確保資料分類可以存取磁碟區。您需要為資料分類提供 CIFS 憑證，以便它可以存取 CIFS 磁碟區。

對於Azure NetApp Files，資料分類只能掃描與控制台位於同一區域的磁碟區。

清單

- 確保資料分類實例與包含Azure NetApp Files區的每個網路之間存在網路連線。
- 確保以下連接埠對資料分類實例開放：
  - 對於 NFS – 連接埠 111 和 2049。
  - 對於 CIFS – 連接埠 139 和 445。
- 確保 NFS 磁碟區匯出策略包含資料分類執行個體的 IP 位址，以便它可以存取每個磁碟區上的資料。

步驟

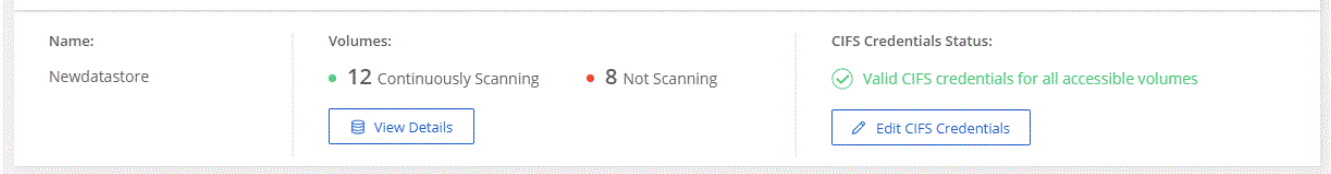
1. 從資料分類選單中，選擇*配置*。

   a. 如果您使用 CIFS（SMB），請確保 Active Directory 憑證正確。對於每個系統，選擇*編輯 CIFS 憑證*，然後輸入資料分類存取系統上的 CIFS 磁碟區所需的使用者名稱和密碼。

      憑證可以是唯讀的；提供管理員憑證可確保資料分類可以讀取任何需要提升權限的資料。憑證儲存在資料分類實例上。

      如果您想要確保檔案的「上次存取時間」不會因資料分類掃描而改變，建議使用者在 CIFS 中具有寫入屬性權限或在 NFS 中具有寫入權限。如果可能，請將 Active Directory 使用者設定為組織中具有所有檔案權限的父群組的一部分。

      輸入憑證後，您應該會看到一條訊息，表示所有 CIFS 磁碟區均已成功驗證。



2. 在設定頁面上，選擇*查看詳細資訊*以查看每個 CIFS 和 NFS 磁碟區的狀態。如有必要，請修正任何錯誤，

例如網路連線問題。

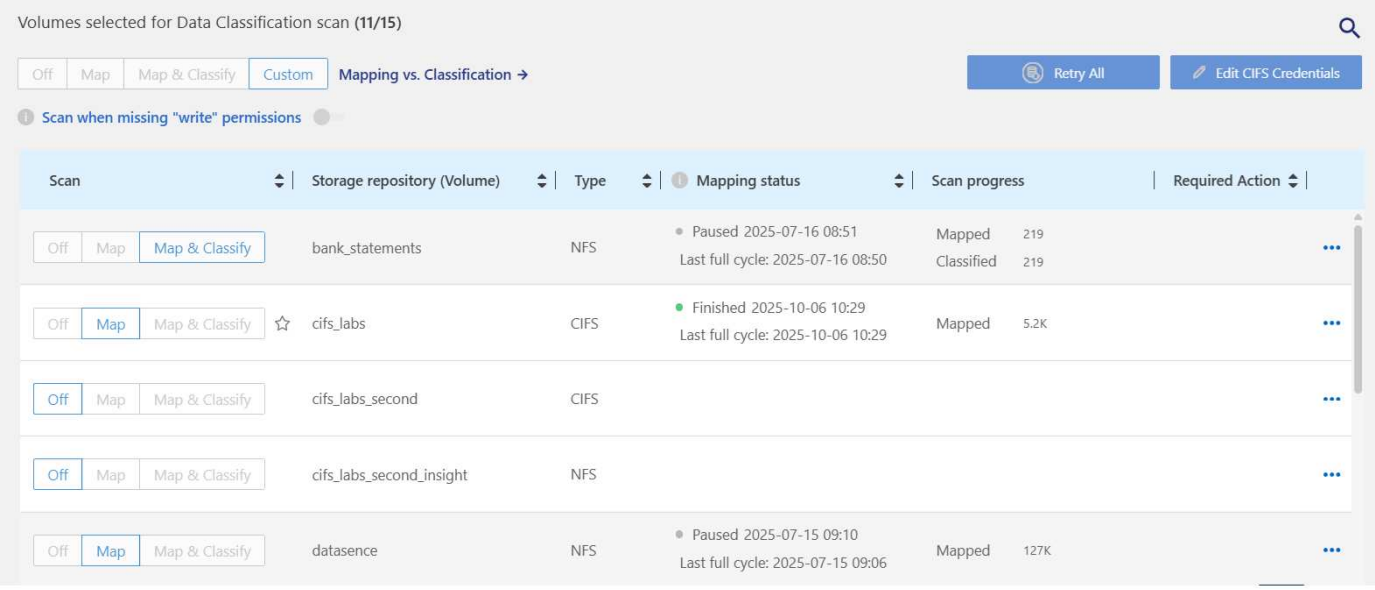
啟用或停用磁碟區掃描

您可以隨時從設定頁面啟動或停止任何系統上的掃描。您也可以將掃描從僅映射掃描切換到映射和分類掃描，反之亦然。建議您掃描系統中的所有磁碟區。

只有當您在標題區域中選擇了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增至系統的新磁碟區。當在標題區域設定為*自訂*或*關閉*時，您需要在系統中新增的每個新磁碟區上啟動對應和/或完整掃描。

頁面頂部的「缺少「寫入」權限時掃描」開關預設為停用狀態。這意味著，如果資料分類在 CIFS 中沒有寫入屬性權限或在 NFS 中沒有寫入權限，系統將不會掃描文件，因為資料分類無法將「上次存取時間」恢復為原始時間戳記。如果您不介意是否重設上次造訪時間，請開啟開關，無論權限為何，都會掃描所有檔案。"了解更多"。

只有當您在標題區域中設定了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增到系統的新磁碟區。當所有磁碟區的設定都是「自訂」或「關閉」時，您需要為新增的每個新磁碟區手動啟動掃描。



步驟

1. 從資料分類選單中，選擇*配置*。
2. 選擇一個系統，然後選擇*配置*。
3. 若要啟用或停用所有磁碟區的掃描，請在所有磁碟區上方的標題中選擇對應、對應和分類或關閉。

   若要啟用或停用對單一卷的掃描，請在清單中找到該卷，然後選擇卷名稱旁的映射、映射和分類或關閉。

結果

當您啟用掃描時，資料分類將開始掃描您在系統中選擇的磁碟區。一旦資料分類開始掃描，結果就會開始出現在合規性儀表板中。掃描完成時間取決於資料量，從幾分鐘到幾小時不等。

使用NetApp Data Classification掃描Cloud Volumes ONTAP和本地ONTAP卷

完成幾個步驟即可開始使用NetApp Data Classification掃描您的Cloud Volumes ONTAP和本機ONTAP磁碟區。

先決條件

在啟用資料分類之前，請確保您具有支援的配置。

- 如果您正在掃描可透過網路存取的Cloud Volumes ONTAP和本機ONTAP系統，您可以"在雲端部署資料分類"或者"在可以存取互聯網的本地位置"。
- 如果您要掃描安裝在沒有網路存取的暗站中的本機ONTAP系統，則需要"在沒有網路存取的相同本地位置部署資料分類"。這要求將控制台代理部署在同一本地位置。

驗證資料分類是否有權存取卷

透過檢查網路、安全群組和匯出策略，確保資料分類可以存取磁碟區。您需要向資料分類提供 CIFS 憑證，以便它可以存取 CIFS 磁碟區。

清單

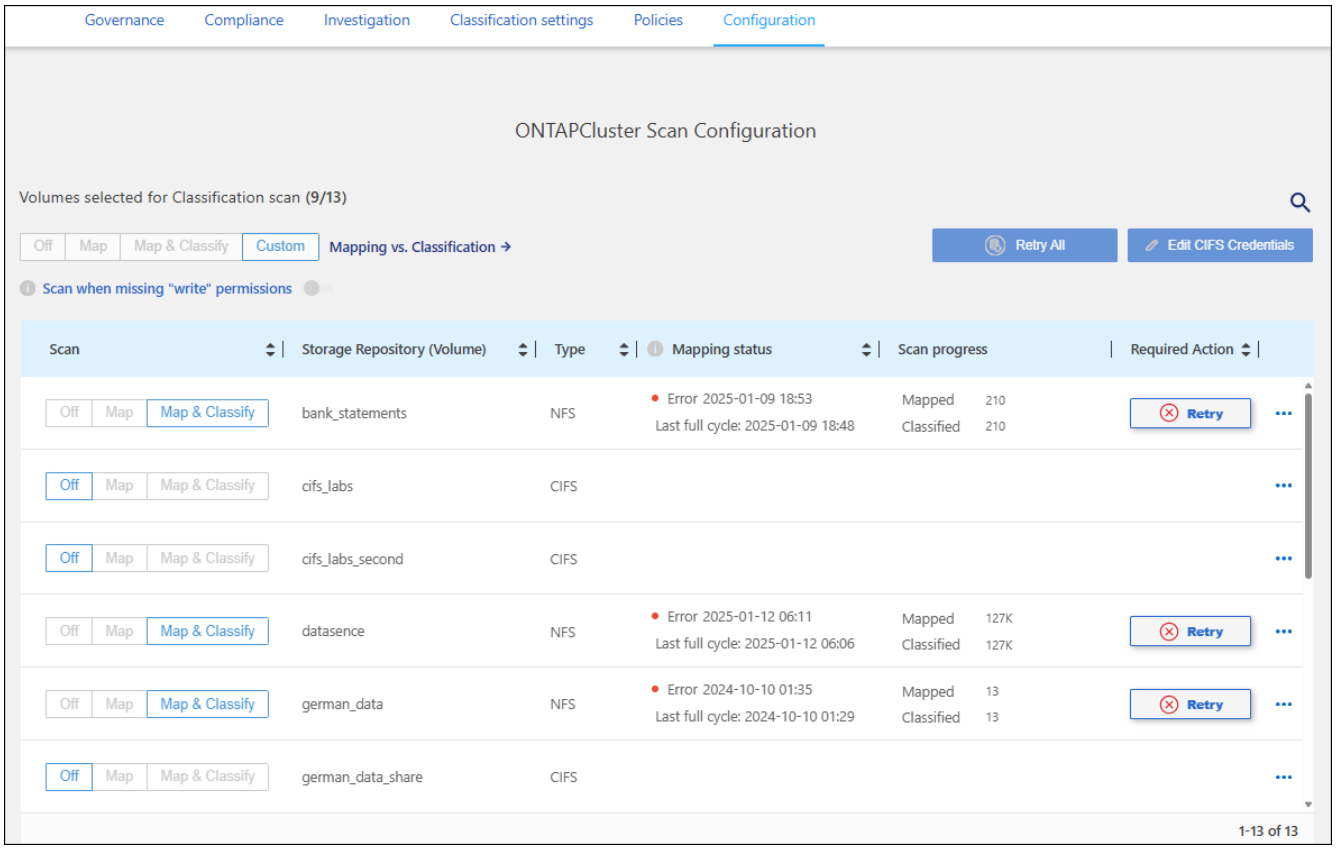
- 確保資料分類實例與包含Cloud Volumes ONTAP或本地ONTAP叢集的磁碟區的每個網路之間存在網路連線。
- 確保Cloud Volumes ONTAP的安全群組允許來自資料分類實例的入站流量。

  您可以為來自資料分類實例的 IP 位址的流量開啟安全性群組，也可以為來自虛擬網路內部的所有流量開啟安全性群組。

- 確保 NFS 磁碟區匯出策略包含資料分類執行個體的 IP 位址，以便它可以存取每個磁碟區上的資料。

步驟

1. 從資料分類選單中，選擇*配置*。



2. 如果您使用 CIFS，請向資料分類提供 Active Directory 憑證，以便它可以掃描 CIFS 磁碟區。對於每個系統，選擇*編輯 CIFS 憑證*並輸入資料分類存取系統上的 CIFS 磁碟區所需的使用者名稱和密碼。

   憑證可以是唯讀的，但提供管理員憑證可確保資料分類可以讀取任何需要提升權限的資料。憑證儲存在資料分類實例上。

   如果您想要確保檔案的「上次存取時間」不會因資料分類掃描而改變，建議使用者在 CIFS 中具有寫入屬性權限或在 NFS 中具有寫入權限。如果可能，請將 Active Directory 使用者設定為組織中具有所有檔案權限的父群組的一部分。

   如果您正確輸入了憑證，則會出現一則訊息確認所有 CIFS 磁碟區均已成功驗證。

3. 在設定頁面上，選擇*設定*以查看每個 CIFS 和 NFS 磁碟區的狀態並修正任何錯誤。

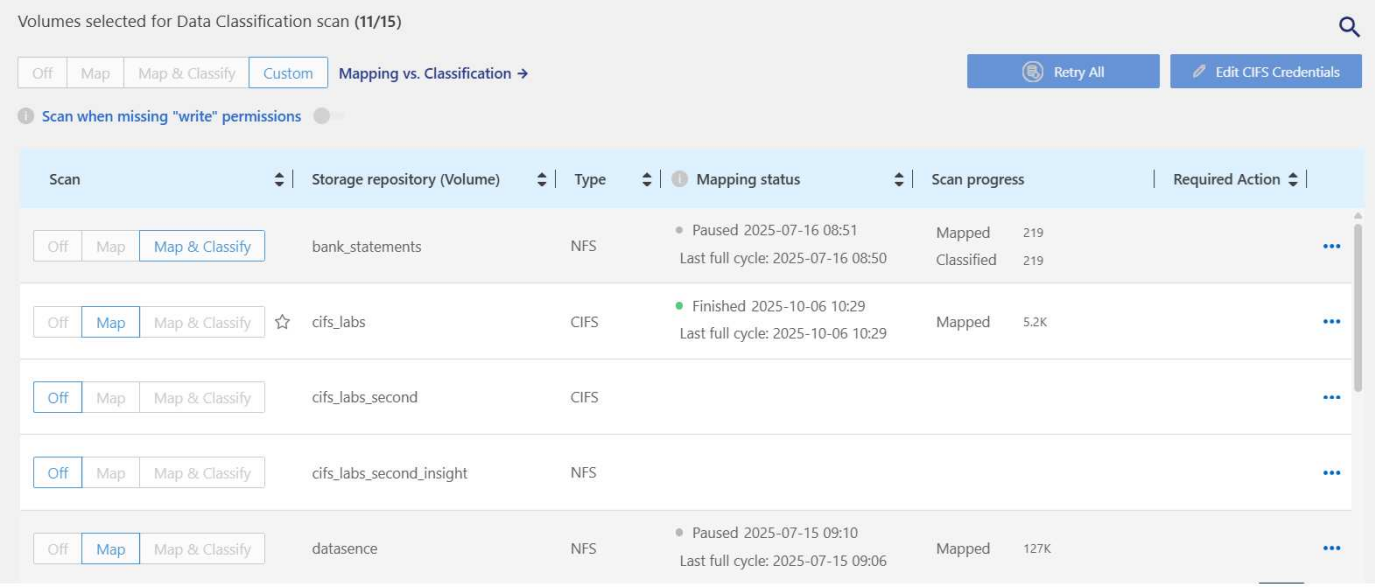
啟用或停用磁碟區掃描

您可以隨時從設定頁面啟動或停止任何系統上的掃描。您也可以將掃描從僅映射掃描切換到映射和分類掃描，反之亦然。建議您掃描系統中的所有磁碟區。

只有當您在標題區域中選擇了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增至系統的新磁碟區。當在標題區域設定為*自訂*或*關閉*時，您需要在系統中新增的每個新磁碟區上啟動對應和/或完整掃描。

頁面頂部的「缺少「寫入」權限時掃描」開關預設為停用狀態。這意味著，如果資料分類在 CIFS 中沒有寫入屬性權限或在 NFS 中沒有寫入權限，系統將不會掃描文件，因為資料分類無法將「上次存取時間」恢復為原始時間戳記。如果您不介意是否重設上次造訪時間，請開啟開關，無論權限為何，都會掃描所有檔案。"了解更多"。

只有當您在標題區域中設定了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增到系統的新磁碟區。當所有磁碟區的設定都是「自訂」或「關閉」時，您需要為新增的每個新磁碟區手動啟動掃描。



步驟

1. 從資料分類選單中，選擇*配置*。
2. 選擇一個系統，然後選擇*配置*。
3. 若要啟用或停用所有磁碟區的掃描，請在所有磁碟區上方的標題中選擇對應、對應和分類或關閉。

   若要啟用或停用對單一卷的掃描，請在清單中找到該卷，然後選擇卷名稱旁的映射、映射和分類或關閉。

結果

當您啟用掃描時，資料分類將開始掃描您在系統中選擇的磁碟區。一旦資料分類開始掃描，結果就會開始出現在合規性儀表板中。掃描完成時間取決於資料量，從幾分鐘到幾小時不等。

資料分類僅掃描卷下的一個檔案共用。如果您的磁碟區中有多個共享，則需要將這些其他共享作為共享群組單獨掃描。"查看有關此數據分類限制的更多詳細信息" 。

使用NetApp Data Classification

完成幾個步驟即可開始使用NetApp Data Classification掃描資料庫模式。

審查先決條件

在啟用資料分類之前，請查看以下先決條件，以確保您具有受支援的配置。

支援的資料庫

資料分類可以掃描以下資料庫中的模式：

- 亞馬遜關係型資料庫服務 (Amazon RDS)

- MongoDB
- MySQL
- 甲骨文
- PostgreSQL
- SAP HANA
- SQL 伺服器 (MSSQL)

資料庫中必須啟用統計資訊收集功能。

資料庫要求

任何與資料分類實例連接的資料庫都可以掃描，無論它託管在何處。您只需要以下資訊即可連接到資料庫：

- IP 位址或主機名
- 港口
- 服務名稱（僅用於存取 Oracle 資料庫）
- 允許讀取架構的憑證

  選擇使用者名稱和密碼時，請務必選擇對要掃描的所有模式和表格具有完全讀取權限的使用者名稱和密碼。我們建議您為資料分類系統建立一個具有所有必需權限的專用使用者。

對於 MongoDB，需要只讀管理員角色。

部署資料分類實例

如果尚未部署實例，則部署資料分類。

如果您正在掃描可透過網際網路存取的資料庫模式，您可以"在雲端部署資料分類"或者"在可以存取網際網路的本地位置部署資料分類"。

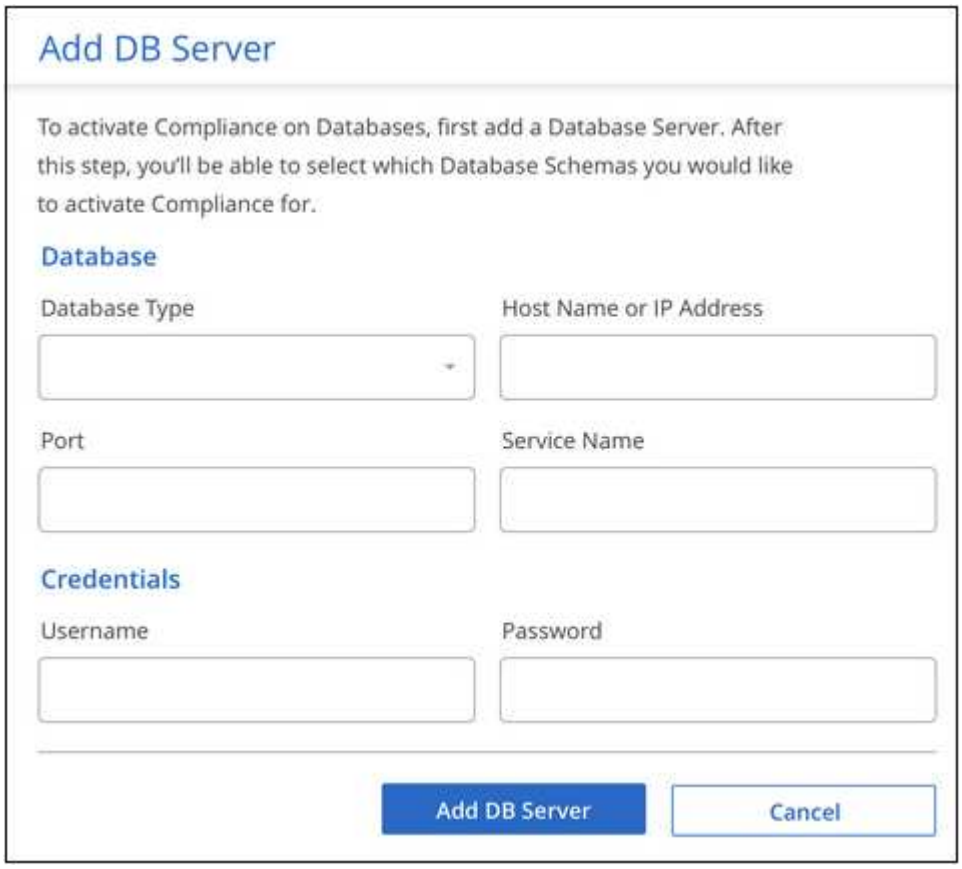
如果您正在掃描安裝在沒有網路存取的暗網中的資料庫模式，則需要"在沒有網路存取的相同本地位置部署資料分類"。這也要求將控制台代理部署在同一本地位置。

新增資料庫伺服器

新增架構所在的資料庫伺服器。

1. 從資料分類選單中，選擇*配置*。
2. 在設定頁面中，選擇*新增系統* > 新增資料庫伺服器。
3. 輸入所需資訊以識別資料庫伺服器。
   a. 選擇資料庫類型。
   b. 輸入連接埠和主機名稱或 IP 位址以連接到資料庫。
   c. 對於 Oracle 資料庫，輸入服務名稱。
   d. 輸入憑證以便資料分類可以存取伺服器。

e. 選擇*新增資料庫伺服器*。



該資料庫已新增至系統清單。

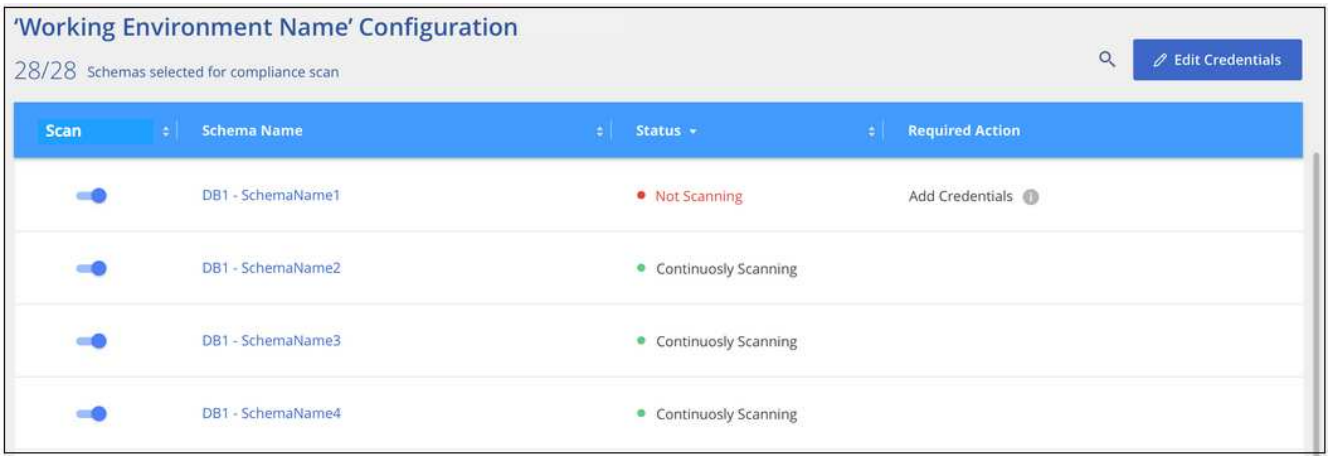
啟用和停用資料庫模式掃描

您可以隨時停止或開始對您的模式進行全面掃描。

沒有選項可以選擇僅映射資料庫模式的掃描。

1. 在配置頁面中，選擇要設定的資料庫的*配置*按鈕。

2. 透過向右移動滑桿來選擇要掃描的模式。



結果

資料分類開始掃描您啟用的資料庫模式。您可以透過導航到配置選單然後選擇系統配置來追蹤初始掃描的進度。每次掃描的進度都顯示為進度條。您也可以將滑鼠停留在進度列上，查看相對於磁碟區中檔案總數的已掃描檔案數。如果有任何錯誤，它們將出現在「狀態」列中，並顯示修復錯誤所需的操作。

資料分類每天掃描您的資料庫一次；資料庫不像其他資料來源那樣被連續掃描。

使用NetApp Data Classification掃描Google Cloud NetApp Volumes

NetApp Data Classification支援Google Cloud NetApp Volumes作為一個系統。了解如何掃描您的Google Cloud NetApp Volumes系統。

發現您要掃描的Google Cloud NetApp Volumes系統

如果您要掃描的Google Cloud NetApp Volumes系統尚未作為系統出現在NetApp Console中，"將其新增至系統頁面" 。

部署資料分類實例

"部署資料分類"如果尚未部署實例。

掃描Google Cloud NetApp Volumes時，資料分類必須部署在雲端中，並且必須部署在與您要掃描的磁碟區相同的區域。

*注意：*掃描Google Cloud NetApp Volumes時，目前不支援在本機位置部署資料分類。

在您的系統中啟用資料分類

您可以在Google Cloud NetApp Volumes系統上啟用資料分類。

1. 從資料分類選單中，選擇*配置*。
2. 選擇如何掃描每個系統中的磁碟區。"了解映射和分類掃描"：
   - 若要對應所有捲，請選擇*映射所有磁碟區*。
   - 若要對應和分類所有捲，請選擇*映射和分類所有捲*。
   - 若要自訂每個磁碟區的掃描，請選擇*或選擇每個磁碟區的掃描類型*，然後選擇要對應和/或分類的磁碟

區。

看啟用和停用磁碟區上的掃描了解詳情。

3. 在確認對話方塊中，選擇*批准*。

結果

資料分類開始掃描您在系統中選擇的磁碟區。一旦資料分類完成初始掃描，結果就會顯示在合規性儀表板中。所需時間取決於資料量：幾分鐘到幾個小時。您可以在配置選單的系統配置部分追蹤初始掃描的進度。資料分類顯示每次掃描的進度條。您也可以將滑鼠停留在進度列上，查看相對於磁碟區中總檔案數的已掃描檔案數。

- 預設情況下，如果資料分類在 CIFS 中沒有寫入屬性權限，或者在 NFS 中沒有寫入權限，系統將不會掃描磁碟區中的文件，因為資料分類無法將「上次存取時間」還原為原始時間戳記。如果您不介意上次造訪時間是否重設，請選擇*或為每個磁碟區選擇掃描類型*。結果頁面有一個您可以啟用的設置，以便資料分類可以掃描卷，而不管權限如何。
- 資料分類僅掃描卷下的一個檔案共用。如果您的磁碟區中有多個共享，則需要將這些其他共享作為共享群組單獨掃描。"了解此資料分類限制"。

驗證資料分類是否有權存取卷

透過檢查您的網路、安全群組和匯出策略，確保資料分類可以存取磁碟區。對於 CIFS 卷，您需要為資料分類提供 CIFS 憑證。

對於Google Cloud NetApp Volumes，資料分類只能掃描與控制台位於同一區域的磁碟區。

清單

- 確保資料分類實例與包含Google Cloud NetApp Volumes的每個網路之間存在網路連線。
- 確保以下連接埠對資料分類實例開放：
  - 對於 NFS – 連接埠 111 和 2049。
  - 對於 CIFS – 連接埠 139 和 445。
- 確保 NFS 磁碟區匯出策略包含資料分類執行個體的 IP 位址，以便它可以存取每個磁碟區上的資料。

步驟

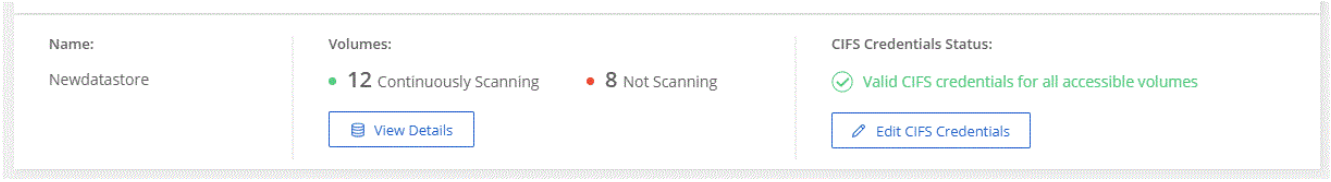
1. 從資料分類選單中，選擇*配置*。

   a. 如果您使用 CIFS（SMB），請確保 Active Directory 憑證正確。對於每個系統，選擇*編輯 CIFS 憑證*，然後輸入資料分類存取系統上的 CIFS 磁碟區所需的使用者名稱和密碼。

      憑證可以是唯讀的，但提供管理員憑證可確保資料分類可以讀取任何需要提升權限的資料。憑證儲存在資料分類實例上。

      如果您想要確保檔案的「上次存取時間」不會因資料分類掃描而改變，建議使用者在 CIFS 中具有寫入屬性權限或在 NFS 中具有寫入權限。如果可能，請將 Active Directory 使用者設定為組織中具有所有檔案權限的父群組的一部分。

      輸入憑證後，您應該會看到一條訊息，表示所有 CIFS 磁碟區均已成功驗證。



2. 在設定頁面上，選擇*查看詳細資訊*以查看每個 CIFS 和 NFS 磁碟區的狀態並修正任何錯誤。

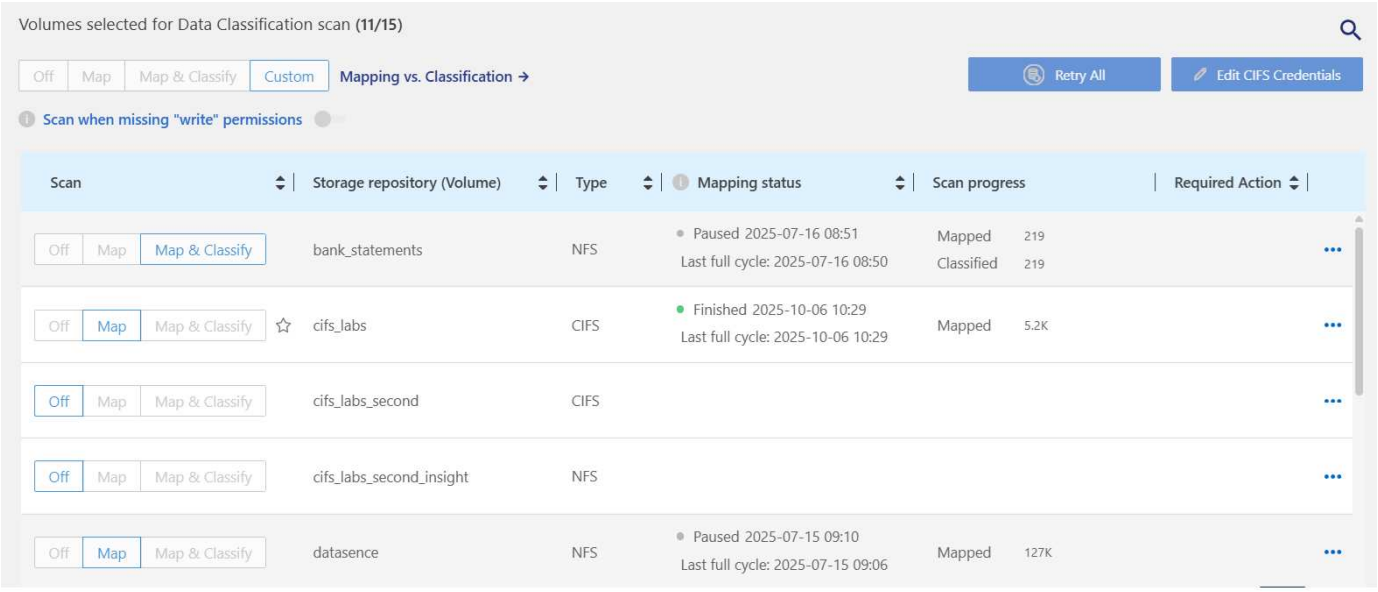
啟用和停用磁碟區上的掃描

您可以隨時從設定頁面啟動或停止任何系統上的掃描。您也可以將掃描從僅映射掃描切換到映射和分類掃描，反之亦然。建議您掃描系統中的所有磁碟區。

只有當您在標題區域中選擇了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增至系統的新磁碟區。當在標題區域設定為*自訂*或*關閉*時，您需要在系統中新增的每個新磁碟區上啟動對應和/或完整掃描。

頁面頂部的「缺少「寫入」權限時掃描」開關預設為停用狀態。這意味著，如果資料分類在 CIFS 中沒有寫入屬性權限或在 NFS 中沒有寫入權限，系統將不會掃描文件，因為資料分類無法將「上次存取時間」恢復為原始時間戳記。如果您不介意是否重設上次造訪時間，請開啟開關，無論權限為何，都會掃描所有檔案。"了解更多"。

只有當您在標題區域中設定了 **Map** 或 **Map & Classify** 設定時，才會自動掃描新增到系統的新磁碟區。當所有磁碟區的設定都是「自訂」或「關閉」時，您需要為新增的每個新磁碟區手動啟動掃描。



步驟

1. 從資料分類選單中，選擇*配置*。
2. 選擇一個系統，然後選擇*配置*。
3. 若要啟用或停用所有磁碟區的掃描，請在所有磁碟區上方的標題中選擇對應、對應和分類或關閉。

   若要啟用或停用對單一卷的掃描，請在清單中找到該卷，然後選擇卷名稱旁的映射、映射和分類或關閉。

結果

當您啟用掃描時，資料分類將開始掃描您在系統中選擇的磁碟區。一旦資料分類開始掃描，結果就會開始出現在合規性儀表板中。掃描完成時間取決於資料量，從幾分鐘到幾小時不等。

使用NetApp Data Classification掃描檔案共用

若要掃描檔案共用，您必須先在NetApp Data Classification中建立檔案共用群組。檔案共用群組適用於本機或雲端中託管的 NFS 或 CIFS (SMB) 共用。

資料分類核心版本不支援掃描非NetApp檔案共享的資料。

先決條件

在啟用資料分類之前，請查看以下先決條件，以確保您具有受支援的配置。

- 共享可以託管在任何地方，包括雲端或本地端。可以將舊版NetApp 7-模式儲存系統中的 CIFS 共用掃描為檔案共用。
  - 資料分類無法從 7 模式系統中提取權限或「上次存取時間」。
  - 由於某些 Linux 版本和 7-模式系統上的 CIFS 共用之間存在已知問題，因此您必須將共用配置為僅使用啟用了 NTLM 驗證的 SMBv1。
- 資料分類實例和共用之間需要有網路連線。
- 您可以將 DFS（分散式檔案系統）共用新增為常規 CIFS 共用。由於資料分類不知道共用是基於組合為單一 CIFS 共用的多個伺服器/磁碟區所建構的，因此當訊息實際上僅適用於位於不同伺服器/磁碟區上的一個資料夾/共用時，您可能會收到有關共用的權限或連線錯誤。
- 對於 CIFS (SMB) 共用，請確保您擁有可提供共用讀取存取權限的 Active Directory 憑證。如果資料分類需要掃描任何需要提升權限的數據，則最好使用管理員憑證。

  如果您想要確保檔案的「上次存取時間」不會因資料分類掃描而改變，建議使用者在 CIFS 中具有寫入屬性權限或在 NFS 中具有寫入權限。如果可能，請將 Active Directory 使用者設定為組織中具有所有檔案權限的父群組的一部分。

- 群組中的所有 CIFS 檔案共用必須使用相同的 Active Directory 憑證。
- 您可以混合使用 NFS 和 CIFS（使用 Kerberos 或 NTLM）共用。您必須單獨將共用新增至群組。也就是說，您必須完成該過程兩次 - 每個協議一次。
  - 您無法建立混合 CIFS 驗證類型（Kerberos 和 NTLM）的檔案共用群組。
- 如果您使用具有 Kerberos 驗證的 CIFS，請確保資料分類可以存取所提供的 IP 位址。如果 IP 位址無法訪問，則無法新增檔案共用。

建立文件共享組

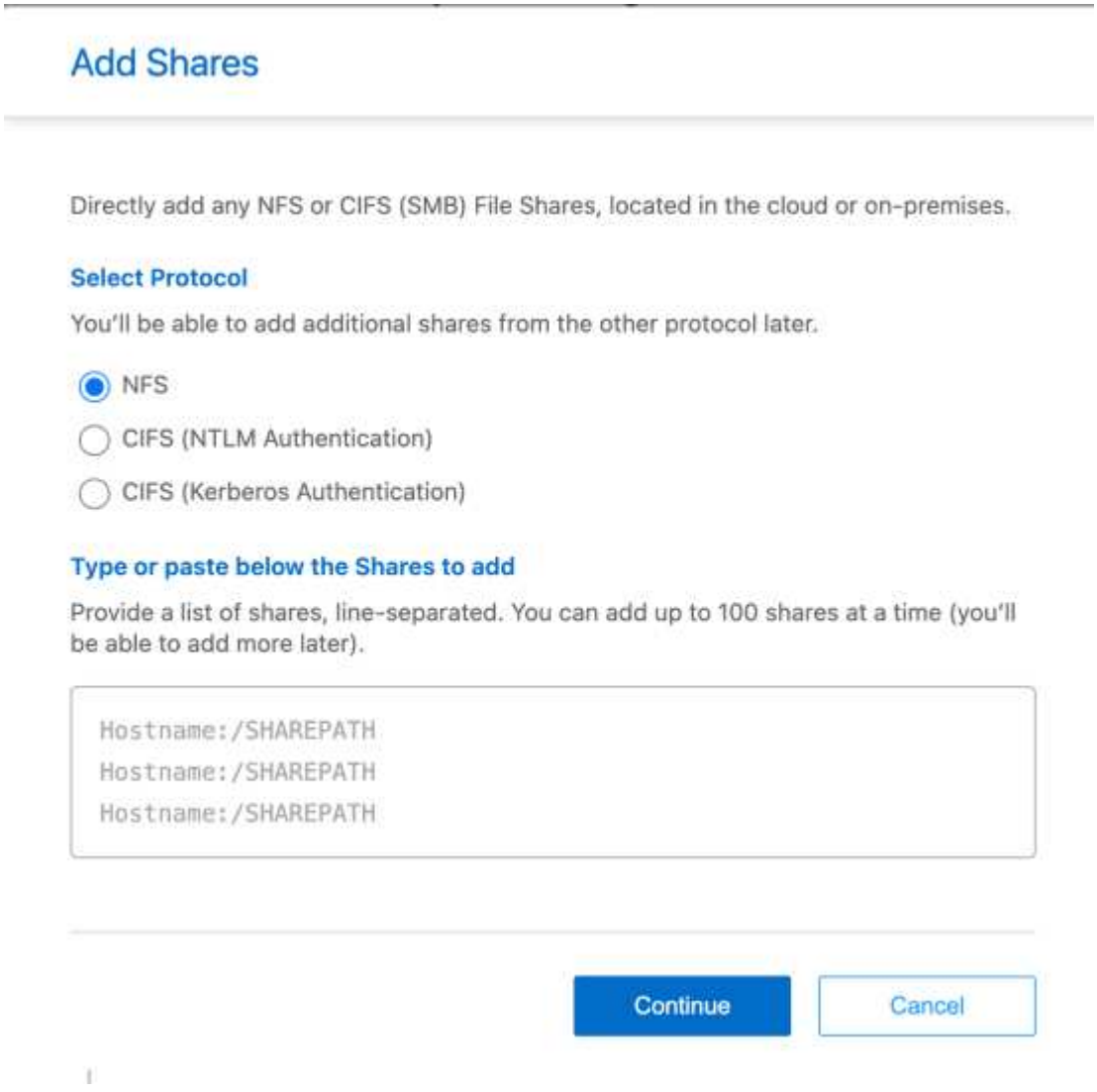
將檔案共用新增至群組時，必須使用格式 `<host_name>:/<share_path>`。

您可以單獨新增文件共享，也可以輸入要掃描的文件共享的行分隔清單。您一次最多可以添加 100 股。

步驟

1. 從資料分類選單中，選擇*配置*。

2. 從設定頁面中，選擇*新增系統*>*新增檔案共用群組*。
3. 在新增檔案共享群組對話方塊中，輸入共享群組的名稱，然後選擇*繼續*。
4. 選擇要新增的文件共享的協定。



   a. 如果您要新增具有 NTLM 驗證的 CIFS 共用，請輸入 Active Directory 憑證以存取 CIFS 磁碟區。儘管支援唯讀憑證，但建議您使用管理員憑證提供完全存取權限。選擇儲存。
5. 新增要掃描的檔案共用（每行一個檔案共用）。然後選擇繼續。
6. 確認對話方塊顯示已新增的共享數量。

   如果對話方塊列出了任何無法新增的共享，請擷取此資訊以便解決問題。如果問題與命名約定有關，您可以使用已修正的名稱重新新增共用。

7. 配置磁碟區上的掃描：
   - 若要對檔案共用啟用僅映射掃描，請選擇*映射*。
   - 若要對檔案共用啟用完整掃描，請選擇*Map & Classify*。
   - 若要停用檔案共用上的掃描，請選擇「關閉」。

頁面頂部的「缺少「寫入屬性」權限時掃描」開關預設為停用狀態。這意味著，如果資料分類在 CIFS 中沒有寫入屬性權限或在 NFS 中沒有寫入權限，系統將不會掃描文件，因為資料分類無法將「上次存取時間」恢復為原始時間戳記。+ 如果將*缺少「寫入屬性」權限時掃描*切換為*開*，則掃描將重置上次存取時間並掃描所有文件，而不管權限如何。+ 要了解有關上次訪問時間戳的更多信息，請參閱"從資料分類中的資料來源收集的元資料"。

結果

資料分類開始掃描您新增的檔案共用中的檔案。你可以追蹤掃描進度並在儀表板中查看掃描結果。

如果對於使用 Kerberos 驗證的 CIFS 配置的掃描未成功完成，請檢查配置選項卡中是否有錯誤。

編輯檔案共享群組

建立檔案共用群組後，您可以編輯 CIFS 協定或新增和刪除檔案共用。

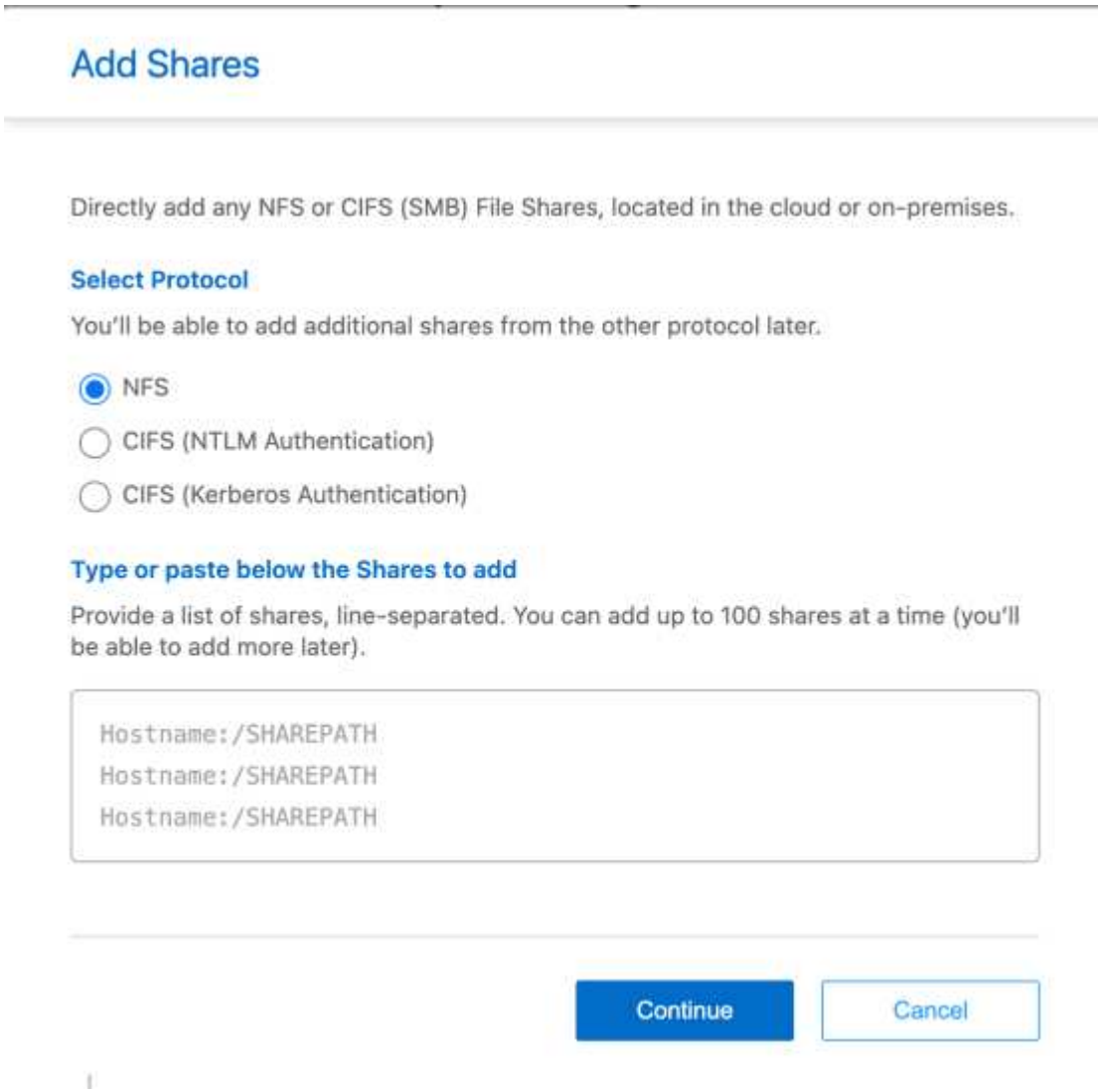
編輯 **CIFS** 協定配置

1. 從資料分類選單中，選擇*配置*。
2. 從設定頁面中，選擇要修改的檔案共用群組。
3. 選擇編輯 **CIFS** 憑證。

4. 選擇身份驗證方法：**NTLM** 或 **Kerberos**。
5. 輸入 Active Directory 使用者名稱和密碼。
6. 選擇儲存以完成該過程。

將文件共享新增至掃描

1. 從資料分類選單中，選擇*配置*。
2. 從設定頁面中，選擇要修改的檔案共用群組。
3. 選擇 **+** 新增共享。
4. 選擇要新增的文件共享的協定。



如果您要將文件共用新增至已設定的協議，則無需進行任何變更。

如果您要使用第二種協定新增檔案共用，請確保您已正確配置身份驗證，詳情請見"先決條件"。

5. 使用以下格式新增要掃描的檔案共用（每行一個檔案共用） `<host_name>:/<share_path>` 。
6. 選擇繼續以完成新增檔案共用。

從掃描中刪除檔案共享

1. 從資料分類選單中，選擇*配置*。
2. 選擇要從中刪除檔案共享的系統。
3. 選擇*配置*。
4. 在配置頁面中，選擇操作 ... 對於要刪除的檔案共用。
5. 從操作選單中，選擇*刪除共用*。

追蹤掃描進度

您可以追蹤初始掃描的進度。

1. 選擇配置選單。
2. 選擇系統配置。
3. 對於儲存庫，檢查掃描進度列以查看其狀態。

使用NetApp Data Classification掃描StorageGRID數據

完成幾個步驟即可直接使用NetApp Data Classification開始掃描StorageGRID內的資料。

查看StorageGRID要求

在啟用資料分類之前，請查看以下先決條件，以確保您具有受支援的配置。

- 您需要有端點 URL 才能連接物件儲存服務。
- 您需要擁有來自StorageGRID的存取金鑰和金鑰，以便資料分類可以存取儲存桶。

部署資料分類實例

如果尚未部署實例，則部署資料分類。

如果您正在掃描可透過網際網路存取的StorageGRID數據，您可以"在雲端部署資料分類"或者"在可以存取網際網路的本地位置部署資料分類"。

如果您要掃描安裝在沒有網路存取的暗站中的StorageGRID數據，則需要"在沒有網路存取的相同本地位置部署資料分類"。這也要求將控制台代理部署在同一本地位置。

將StorageGRID服務新增至資料分類

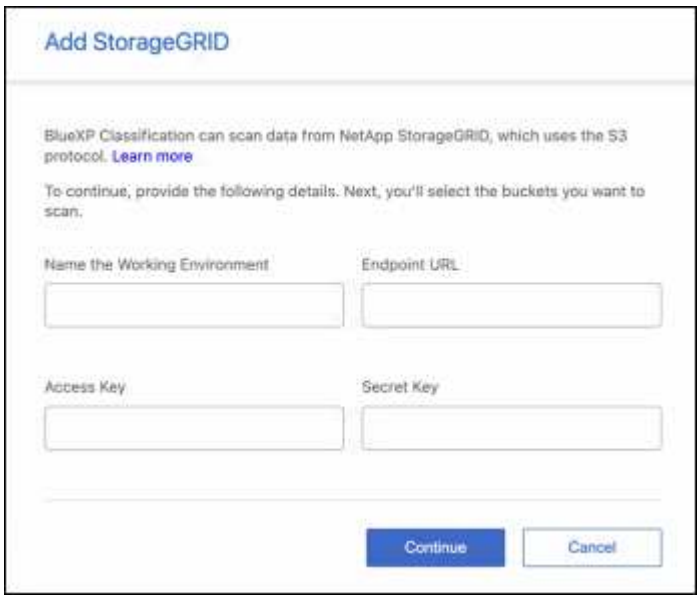
新增StorageGRID服務。

步驟

1. 從資料分類選單中，選擇*配置*選項。
2. 在設定頁面中，選擇「新增系統」>「新增StorageGRID」。
3. 在新增StorageGRID服務對話方塊中，輸入StorageGRID服務的詳細資訊並選擇*繼續*。
    a. 輸入您想要使用的系統名稱。此名稱應反映您要連線的StorageGRID服務的名稱。

b. 輸入 Endpoint URL 以存取物件儲存服務。

c. 輸入存取密鑰和密鑰，以便資料分類可以存取StorageGRID中的儲存桶。



結果

StorageGRID已新增至系統清單。

啟用並停用**StorageGRID**桶上的掃描

在StorageGRID上啟用資料分類後，下一步是設定要掃描的儲存桶。資料分類發現這些儲存桶並將它們顯示在您建立的系統中。

步驟

1. 在設定頁面中，找到StorageGRID系統。
2. 在StorageGRID系統圖塊上，選擇 配置。
3. 完成以下步驟之一來啟用或停用掃描：
   - 若要對儲存體桶啟用僅對應掃描，請選擇*Map*。
   - 若要對儲存桶啟用完整掃描，請選擇*Map & Classify*。
   - 若要停用儲存桶的掃描，請選擇「關閉」。

結果

資料分類開始掃描您啟用的儲存桶。您可以透過導航到配置選單然後選擇系統配置來追蹤初始掃描的進度。每次掃描的進度都顯示為進度條。您也可以將滑鼠停留在進度列上，查看相對於磁碟區中總檔案數的已掃描檔案數。如果有任何錯誤，它們將出現在「狀態」列中，並顯示修復錯誤所需的操作。

將您的 **Active Directory** 與**NetApp Data Classification**集成

您可以將全域 Active Directory 與NetApp Data Classification集成，以增強資料分類報告有關文件擁有者以及哪些使用者和群組有權存取您的文件的結果。

當您設定某些資料來源（如下所列）時，您需要輸入 Active Directory 憑證以便資料分類掃描 CIFS 磁碟區。此集成為資料分類提供了駐留在這些資料來源中的資料的檔案擁有者和權限詳細資訊。為這些資料來源輸入的 Active Directory 可能與您在此輸入的全域 Active Directory 憑證不同。資料分類將在所有整合的活動目錄中尋找使用者和權限詳細資訊。

此整合在資料分類的以下位置提供了附加資訊：

- 您可以使用“文件所有者”"篩選"並在調查窗格中的文件元資料中查看結果。它不是包含 SID（安全性識別碼）的檔案擁有者，而是填入實際的使用者名稱。

  您也可以查看更多關於檔案擁有者的詳細資訊：帳戶名稱、電子郵件地址和 SAM 帳戶名稱，或查看該使用者擁有的項目。

- 你可以看到"完整檔案權限"當您按一下「查看所有權限」按鈕時，每個檔案和目錄都會顯示所有權限。
- 在"治理儀表板"，開啟權限面板將顯示有關您的資料的更詳細的詳細資訊。

本機使用者 SID 和來自未知網域的 SID 不會轉換為實際使用者名稱。

支援的資料來源

Active Directory 與資料分類的整合可以識別來自下列資料來源的資料：

- 本地ONTAP系統
- Cloud Volumes ONTAP
- Azure NetApp Files
- 適用於ONTAP的 FSx

連接到您的 **Active Directory** 伺服器

部署資料分類並啟動資料來源掃描後，您可以將資料分類與 Active Directory 整合。可以使用 DNS 伺服器 IP 位址或 LDAP 伺服器 IP 位址存取 Active Directory。

Active Directory 憑證可以是唯讀的，但提供管理員憑證可確保資料分類可以讀取任何需要提升權限的資料。憑證儲存在資料分類實例上。

對於 CIFS 磁碟區/檔案共用，如果您想確保檔案的「上次存取時間」不會因資料分類掃描而改變，則使用者應該具有寫入屬性權限。如果可能的話，我們建議將 Active Directory 設定的使用者作為組織中對所有檔案具有權限的父群組的一部分。

要求

- 您必須已經為公司使用者設定了 Active Directory。
- 您必須具有 Active Directory 的資訊：

  - DNS 伺服器 IP 位址，或多個 IP 位址

    或者

  LDAP 伺服器 IP 位址，或多個 IP 位址

- 存取伺服器的使用者名稱和密碼
- 網域名稱（Active Directory 名稱）
- 您是否使用安全 LDAP (LDAPS)
- LDAP 伺服器連接埠（LDAP 通常為 389，安全 LDAP 通常為 636）

- 必須開啟以下連接埠以供資料分類實例進行出站通訊：

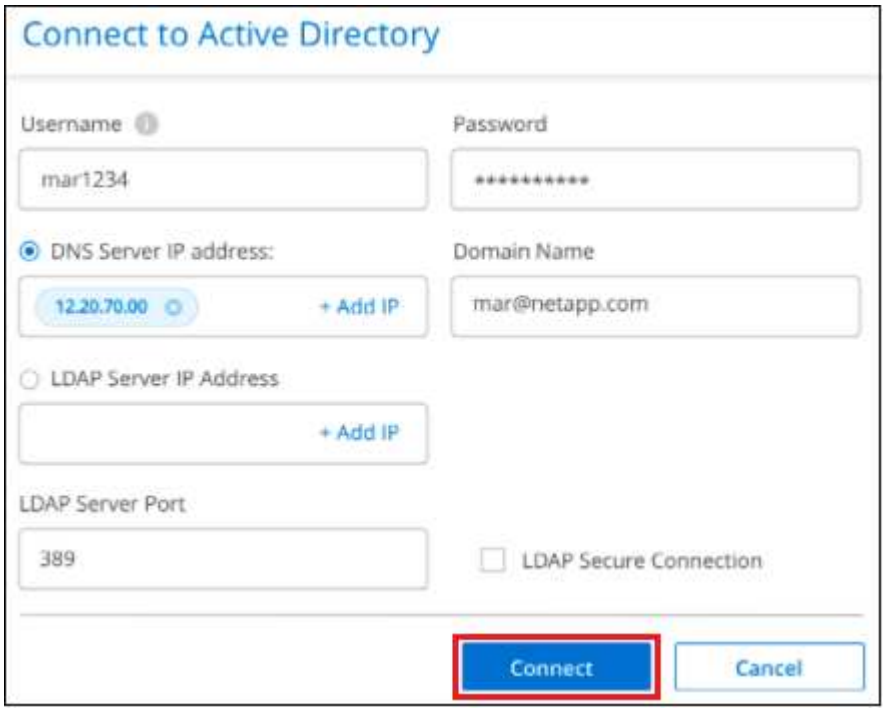
協定	港口	目的地	目的
TCP 和 UDP	389	活動目錄	LDAP
TCP	636	活動目錄	透過 SSL 的 LDAP
TCP	3268	活動目錄	全域目錄
TCP	3269	活動目錄	透過 SSL 的全域目錄

步驟

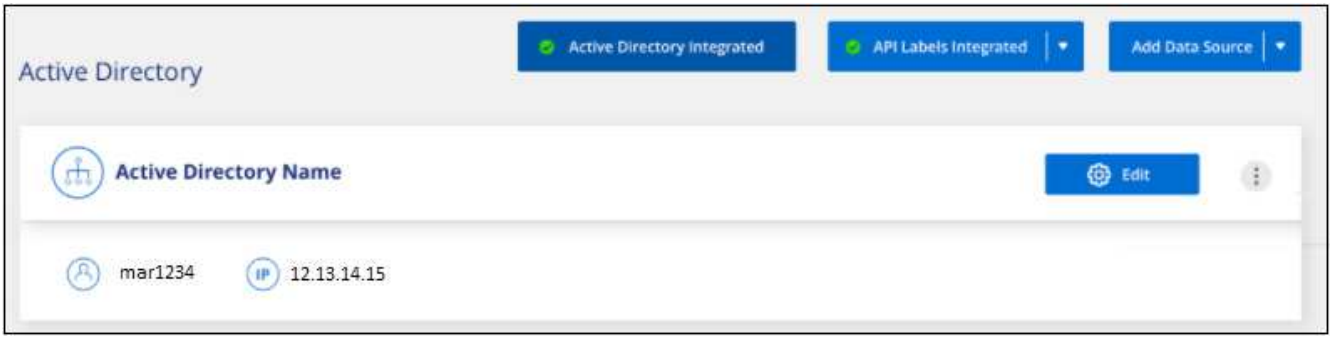
1. 在資料分類設定頁面中，按一下*新增 Active Directory*。

2. 在「連接到 Active Directory」對話方塊中，輸入 Active Directory 詳細訊息，然後按一下「連線」。

   如果需要，您可以透過選擇「新增 IP」來新增多個 IP 位址。



   資料分類整合到 Active Directory，並在設定頁面中新增一個新部分。



管理您的 **Active Directory** 集成

如果您需要修改 Active Directory 整合中的任何值，請按一下「編輯」按鈕並進行變更。

您也可以選擇 按鈕，然後*刪除 Active Directory*。